

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 3, 2003, 23:31:53 ; Search time 139 Seconds  
(without alignments)  
3290.435 Million cell updates/sec

Title: US-10-062-624-40  
Perfect score: 1496  
Sequence: 1 MNKKLKEITNTVLVCLLSL.....AVATLNTGYGGEIGARLTF 293

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10062624/runat\_30062003\_091231\_24413/app\_query.fasta\_1.455  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCI=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10062624 @CGN 1 1 118 @runat\_30062003\_091231\_24413  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:\*\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq\*\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq\*\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq\*\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq\*\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq\*\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq\*\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq\*\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq\*\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq\*\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq\*\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq\*\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq\*\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq\*\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	1496	100.0	879	9	US-10-062-051-39 Sequence 39, Appl
2	1496	100.0	879	9	US-10-062-920-39 Sequence 39, Appl
3	1496	100.0	879	12	US-10-062-624-39 Sequence 39, Appl
4	1483	99.1	882	9	US-10-314-639-43 Sequence 43, Appl

#### SUMMARIES

US-10-062-051-39  
Alignment Scores: 2.58e-164 Length: 879  
Pred. No.: 879

#### ALIGNMENTS

##### RESULT 1

US-10-062-051-39  
; Sequence 39, Application US/10062051  
; Publication No. US20030073095A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: 2002-01-31  
; PRIOR FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: US/09/660,587  
; PRIOR FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 39  
; LENGTH: 879  
; TYPE: DNA  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: nucleic acid sequence of E. canis p28-1

Sequence 43, Appl  
Sequence 17, Appl  
Sequence 17, Appl  
Sequence 13, Appl  
Sequence 13, Appl  
Sequence 55, Appl  
Sequence 55, Appl  
Sequence 51, Appl  
Sequence 51, Appl  
Sequence 49, Appl  
Sequence 49, Appl  
Sequence 45, Appl  
Sequence 45, Appl  
Sequence 59, Appl  
Sequence 59, Appl  
Sequence 23, Appl  
Sequence 23, Appl  
Sequence 29, Appl  
Sequence 29, Appl  
Sequence 25, Appl  
Sequence 25, Appl  
Sequence 57, Appl  
Sequence 57, Appl  
Sequence 27, Appl  
Sequence 27, Appl  
Sequence 53, Appl  
Sequence 53, Appl  
Sequence 21, Appl  
Sequence 21, Appl  
Sequence 61, Appl  
Sequence 61, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 41, Appl  
Sequence 41, Appl  
Sequence 47, Appl  
Sequence 47, Appl  
Sequence 11, Appl  
Sequence 11, Appl  
Sequence 31, Appl

Score:	1496.00	Matches:	293
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0
US-10-062-624-40 (1-293) x US-10-062-051-39 (1-879)			
Qy	1	MetAsnAsnLysLeuLysPheThrIleIleAsnThrValLeuValCysLeuSerLeu	20
Db	1	ATGAATAATAAACTCAAAATTTACTATATAATAACACAGTATTAGTAGCTTTATTGTCAATTA	60
Qy	21	ProAsnIleSerSerSerLysAlaIleAsnAsnAlaLysLysTyTyTyGlyLeuTy	40
Db	61	CCTAAATATATCTTCTCAAGGCCATAAACCAATACGCTAAAAGTAGTACGGATTATAT	120
Qy	41	IleSerGlyGlnTyLysProSerValSerValPheSerAsnPheSerValLysGluThr	60
Db	121	ATCAGTGGCAATATATAACCCAGTGTTCTGTCTTCAGTAATTTTTCAGTTAAAGAAAC	180
Qy	61	AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys	80
Db	181	AATGTCAATAACTAAAAACCTTATAGCTTTAAAAAAGATGTGCACCTATTATGAACCAAG	240
Qy	81	ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyThrAlaVal	100
Db	241	ACTGATGCCAGGTAGGTATTAGTAAACCCATCAAAATTTTACTATCCCTATACAGCTGTA	300
Qy	101	PheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyThrPheAlaGluGlyThr	120
Db	301	TTTCAAGATAAATCTGTCAATTTCAATGGAACTATTGGTTACACCTTTGCTGAAGGTACA	360
Qy	121	ArgValGluIleGluGlySerTyGluGluPheAspValLysAsnProGlyGlyTyThr	140
Db	361	AGAGTTGAAATAGAAGGTTCTTATGAGGAAATTTGATGTAAAAACCCCTGGAGGCTATACA	420
Qy	141	LeuSerAspAlaTyArgTyPheAlaLeuAlaargGluMetLysGlyAsnSerPheThr	160
Db	421	CTAAGTGATGCCCTATCGCTATTTTTGCATTAGCAGTGAATGAAAGGTAAATAGTTTACA	480
Qy	161	ProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSer	180
Db	481	CCTAAGAAAAAAGTTTCTTAATAGTATTTTTTCCACACTGTAAATGAGAAATGATGCATTATCT	540
Qy	181	IleIleSerValIleValAsnValCysTyAspPheSerLeuAsnAsnLeuSerIleSer	200
Db	541	ATAATATCTGTTATAGTAAATGTTTGGCTACAGATTCTCTTTTGAAACAAATTTGTCAATATCG	600
Qy	201	ProTyIleCysGlyGlyValaGlyValAspAlaIleGluPhePheAspValLeuHisIle	220
Db	601	CCTTACATATGTGGAGGACGGGTAGATGCTATAGNAATTCCTCGATGTATTACACATT	660
Qy	221	LysPheAlaTyGlnSerLysLeuGlyIleAlaTySerLeuProSerAsnIleSerLeu	240
Db	661	AAAGTTGCATATCAAAAGCAAGCTAGGTATTGCTATTCTCTTACCATCTAACATTAGTCTC	720
Qy	241	PheAlaSerLeuTyTyHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGln	260
Db	721	TTTGCTAGTGTATATATACCATAAAGTAATGGCCAAATCAATTTAAAAAATTTAAATGTC	780
Qy	261	HisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIle	280
Db	781	CATGTTGCTGAACCTTGAAGTATACCTAAATTTACATCCGAGTTGTACACTTAATATT	840
Qy	281	GlyTyPheGlyGlyGluIleGlyAlaArgLeuThrPhe	293
Db	841	GGTTATTTTGGAGGTGAAATTTGGTCAAGATTGACATTT	879

```

; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/10/062,920
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US/09/660,587
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 05/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 39
; LENGTH: 879
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: Nucleic acid sequence of E. canis p28-1
US-10-062-920-39

Alignment Scores:
Pred. No.:      2,58e-164      Length:      879
Score:          1496.00      Matches:      293
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels:      0
DB:              Gaps:      0

US-10-062-624-40 (1-293) x US-10-062-920-39 (1-879)
Qy      1  MetAsnAsnLysLeuLysPheThrIleIleAsnThrValLeuValCysLeuLeuSerLeu 20
Db      1  ATGATTAATAAAGCTCAAAATTTACTATATAAACACAGATTAGTATGCTTATTGTCATTA 60
Qy     21  ProAsnIleSerSerLysAlaIleAsnAsnAsnAlaLysLysTyTyTyrGlyLeuTyr 40
Db     61  CCTAATAATATCTTCTCTCAAGGCCATAAACAATAACGCTAAAGAGTACTACGGATTATAT 120
Qy     41  IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60
Db    121  ATCAGTGGACAAATATAAACCCAGGTGTTCTGTTTTCAGTAATTTTTTCAGTTAAGAAACC 180
Qy     61  AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
Db    181  AATGTCATACTAAAGAACCTATAGCTTTTAAAGAGATGTTGACTCTATTGAAACCAAG 240
Qy     81  ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaVal 100
Db    241  ACTGATGCCAGTGAGGTATTAGTAACCCATCAAAATTTTACTATCCCTATACAGCTGTA 300
Qy    101  PheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThr 120
Db    301  TTTCAAGATAAATTCGTCAATTTTCAATGGAAGTATTGGTTTACACCTTTGCTGAAGGTACA 360
Qy    121  ArgValGluIleGluGlySerTyrGluPheAspValLysAsnProGlyGlyTyrThr 140
Db    361  AGATTGAAATAGAAGGTCTTATAGGAATTTGATGTTTAAAGCCCTGGAGGCTATACA 420
Qy    141  LeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThr 160
Db    421  CTAAGTATGCTATCGCTATTTTGGCATTTAGCACGCTGAAATGAAAGTAATAGTTTACA 480
Qy    161  ProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSer 180
Db    481  CCTAAAGAAAAGTTCTCTAATAGTATTTTTTCACACTGTGAATGAGAAATGATGGATTATCT 540
Qy    181  IleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSer 200
Db    541  ATAATATCTGTTATAGTAATAATGTTTGCTACGATTTCTTTTGAAACAATTTGTCAATATCG 600
Qy    201  ProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIle 220

```



Db 601 CCTACATATGTGGAGGACAGGGGTAGATGCTATAGAAATTCCTCGATGATTAACACATT 660  
 Qy 221 LysPheAlaTyrGlnSerLysLeuGlyLeuAlaTyrSerLeuProSerAsnIleSerLeu 240  
 Db 661 AAGTTTGCATATCAAGCAAGCTAGGTATGCTTATCTCTACCATCTCAACATTAGTCTC 720  
 Qy 241 PheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGln 260  
 Db 721 TTTGCTAGTTATATTAACATAAAGTAATGGCAATCAATTTAAAAAATTTAAATGTCCAA 780  
 Qy 261 HisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIle 280  
 Db 781 CATGTTGCTGAACCTTGCAGATGATACCTAAATTAACATCCGAGTGTGTACACTTAATATT 840  
 Qy 281 GlyTyrPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293  
 Db 841 GGTATTGAGGTGAAATTTGGTGAAGATTGACATTT 879

RESULT 3  
 US-10-062-624-39  
 ; Sequence 39, Application US/10062624  
 ; Patent No. US20020115840A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Walker, David H.  
 ; APPLICANT: McBride, Jere W.  
 ; APPLICANT: Yu, Xue-Jie  
 ; TITLE OF INVENTION: Homologous 28-Kilodalton Immunodominant Protein  
 ; FILE REFERENCE: D6152CIP2/D1  
 ; CURRENT APPLICATION NUMBER: US/10/062,624  
 ; CURRENT FILING DATE: 2002-01-31  
 ; PRIOR APPLICATION NUMBER: 09/660,587  
 ; PRIOR FILING DATE: 2000-09-12  
 ; NUMBER OF SEQ ID NOS: 46  
 ; SEQ ID NO 39  
 ; LENGTH: 879  
 ; TYPE: DNA  
 ; ORGANISM: Ehrlichia canis  
 ; FEATURE:  
 ; OTHER INFORMATION: nucleic acid sequence of E. canis p28-1

US-10-062-624-39

Alignment Scores:  
 Pred. No.: 2,58e-164 Length: 879  
 Score: 1496.00 Matches: 293  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 12 Gaps: 0

US-10-062-624-40 (1-293) x US-10-062-624-39 (1-879)

Qy 1 MetAsnLysLeuLysPheThrIleLeuAsnThrValLeuValCysLeuLeuSerLeu 20  
 Db 1 ATGAATAATAACCTAAATTTACTATATAATAACACAGTATTAGTATGCTTATTGCTATTA 60  
 Qy 21 ProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40  
 Db 61 CCTAATATATCTTCTCAAGGCCATAAACATTAACGCTAAAGTACTACGATTTATAT 120  
 Qy 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60  
 Db 121 ATCAGTGACAATATAAACCCAGGTGTTCTGTTTTCAGTAATTTTTCAGTTAAAGAACCC 180  
 Qy 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80  
 Db 181 AATGTCATAAATAACCTTATAGCTTTAAAAAAGATGTTGACTCTATTGAAACCAAG 240  
 Qy 81 ThrAspAlaSerValGlyLysSerAsnProSerAsnPheThrIleProTyrThrAlaVal 100  
 Db 241 ACTGATGCCAGTAGGTAGTATTAGTAACCATCAAAATTTTACTATCCCTATACAGCTGTA 300  
 Qy 101 PheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThr 120

Db 301 TTTCAAGATAATTCGTCAATTTCAATGGAACCTATTGGTTACACCTTTGCTGAAGGTACA 360  
 Qy 121 ArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThr 140  
 Db 361 AGAGTTGAAATAGAAAGTTCTTATGAGGAATTTGATGTTAAAAACCTCGAGGCTATACA 420  
 Qy 141 LeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThr 160  
 Db 421 CTAAGTCATGCTATCCTATTTTTCATGCTAGCAGCTGAAATGAAAGGTAAATAGTTTTTACA 480  
 Qy 161 ProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSer 180  
 Db 481 CCTAAAGAAAAAGTTTCTAATAGTATTTTTCACACTCTAATGAGAAATGATGGAATTTCT 540  
 Qy 181 IleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnLeuSerIleSer 200  
 Db 541 ATAATACTCTTATAGTAATGTTTGTACGATTTCTTTTGAACAATTTTGTCAATATCG 600  
 Qy 201 ProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIle 220  
 Db 601 CCTTACATATGTGGAGGAGCAGGGGTAGATGCTATAGAAATTTCTTCGATGTATTACACATT 660  
 Qy 221 LysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeu 240  
 Db 661 AAGTTTGCATATCAAGCAAGCTAGGTATGCTTATCTCTACCATCTCAACATTAGTCTC 720  
 Qy 241 PheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGln 260  
 Db 721 TTTGCTAGTTATATTAACCATAAAGTAATGGCAATCAATTTAAAAAATTTAAATGTCCAA 780  
 Qy 261 HisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIle 280  
 Db 781 CATGTTGCTGAACCTTGCAGATGATACCTAAATTAACATCCGAGTGTGTACACTTAATATT 840  
 Qy 281 GlyTyrPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293  
 Db 841 GGTATTGAGGTGAAATTTGGTGAAGATTGACATTT 879

RESULT 4  
 US-10-314-639-43  
 ; Sequence 43, Application US/10314639  
 ; Publication No. US20030103991A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rikihisa, Yasuko  
 ; APPLICANT: Ohasi, No. US20030103991A1  
 ; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
 ; FILE REFERENCE: Chafensis  
 ; FILE REFERENCE: 22727/04021  
 ; CURRENT APPLICATION NUMBER: US/10/314,639  
 ; CURRENT FILING DATE: 2002-12-09  
 ; PRIOR APPLICATION NUMBER: US/09/314,701  
 ; PRIOR FILING DATE: 1999-05-19  
 ; NUMBER OF SEQ ID NOS: 66  
 ; SOFTWARE: Patent In Ver. 2.0  
 ; SEQ ID NO 43  
 ; LENGTH: 882  
 ; TYPE: DNA  
 ; ORGANISM: Ehrlichia canis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)...(882)  
 US-10-314-639-43

Alignment Scores:  
 Pred. No.: 8,43e-163 Length: 882  
 Score: 1483.00 Matches: 291  
 Percent Similarity: 99.32% Conservative: 0  
 Best Local Similarity: 99.32% Mismatches: 2  
 Query Match: 99.13% Indels: 0  
 DB: 9 Gaps: 0

US-10-062-624-40 (1-293) x US-10-314-639-43 (1-882)

```
QY 1 MetAenAenLysLeuLysPheThrIleAenThrValLeuValCysLeuLeuSerLeu 20
Db 1 ATGATAATAAACTCAATTTACTATAATAAACACAGTATTAGTAGTCTTATTGTCAATTA 60
QY 21 ProAenIleSerSerLysValAlaIleAenAenAenAlaLysLysTyrTyrGlyLeuTyr 40
Db 61 CCTAATATATCTCTCAAGAGCCATAACAAATTAACGCTAAAGTAAGTACGAGTATATAT 120
QY 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAenPheSerValLysGluThr 60
Db 121 ATCAGTGACCAATATAAACCCAGTGTCTCTGTTTTCAGTAATTTTCAGTTAAAGAAC 180
QY 61 AenValIleThrLysAenLeuIleAlaLysLysLysAenValSerValSerIleGluThrLys 80
Db 181 AATGTCAATACTAAACCCCTTATAGCTTTTAAAGAGATGTTGACTCTATTGAAACCAAG 240
QY 81 ThrAenAlaSerValGlyLysSerAenProSerAenPheThrIleProTyrThrAlaVal 100
Db 241 ACTGATGCCAGTGTAGGTATTAGTAACCCATCAAAATTTTACTATCCCTATACAGCTGTA 300
QY 101 PheGlnAenAenSerValAenPheAenGlyThrIleGlyTyrThrPheAlaGluGlyThr 120
Db 301 TTTCAAGATAATCTGTCAATTTCAATGGAATCTTGGTTACACCTTTGCTGAAGGTACA 360
QY 121 ArgValGluIleGluGlySerTyrGluGluPheAenValLysAenProGlyGlyTyrThr 140
Db 361 AGAGTTGAATAGAAAGTCTTATGAGGAATTTGATGTTTAAAGAACCCCTGGAGGCTATACA 420
QY 141 LeuSerAenAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAenSerPheThr 160
Db 421 CTAAAGTGATGCCATCTGCTATTTGCAATAGCAGCTGAATGAAAGGTAATAGTTTACA 480
QY 161 ProLysGluLysValSerAenSerIlePheHisThrValMetArgAenAenProGlyLeuSer 180
Db 481 CCTAAGAAAGAAAGTTTCTAATAGTTTCTTCACTGTAATCAGAAATGATGATTAATCT 540
QY 181 IleIleSerValIleValAenValCysTyrAenPheSerLeuAenAenSerIleSer 200
Db 541 ATAATATCTGTTATAGTAATCTTGTGCTACGATTTCTCTTTGAACAAATTTGTCAATATCG 600
QY 201 ProTyrIleCysGlyValAlaGlyValAenAlaIleGluPheAenValLeuHisIle 220
Db 601 CCTACATATGTGGAGGAGCGGTAGTGTATAGTAATCTTCAATGATTTATACACATT 660
QY 221 LysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAenIleSerLeu 240
Db 661 AAGTTTGCATATCAAAAGCAAGCTAGTATTCTTCTTCACTTACCTTCACTTAAATATT 720
QY 241 PheAlaSerLeuTyrTyrHisLysValMetGlyAenGlnPheLysAenLeuAenValGln 260
Db 721 TTTGCTAGTTTATATACCAATAGTAATCTTAAAGTAAATCAATTTTAAAGTATCCAA 780
QY 261 HisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAenIle 280
Db 781 GATGTTGCTGAACCTTGCAAGTATACCTAATAATTTACATCCGCAATGCTTCACTTAAAT 840
QY 281 GlyTyrPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293
Db 841 GCTTATTTTGGAGGTGAATTTGGTGAAGATTTGACATTT 879
```

## RESULT 5

```
US-10-059-964-43
; Sequence 43, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Bihisa, Yasuko
; APPLICANT: Chasi, No. US20020120115A1
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; CURRENT FILING DATE: 2002-01-28
```

```
;; EARLIER APPLICATION NUMBER: 09/314,701
;; EARLIER FILING DATE: 1999-05-19
;; NUMBER OF SEQ ID NOS: 66
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 43
```

```
;; LENGTH: 882
;; TYPE: DNA
;; ORGANISM: Ehrlichia canis
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(882)
US-10-059-964-43
```

## Alignment Scores:

```
Pred. No.: 8 43e-163 Length: 882
Score: 1483.00 Matches: 291
Percent Similarity: 99.32% Conservative: 0
Best Local Similarity: 99.32% Mismatch: 2
Query Match: 99.13% Indels: 0
DB: 12 Gaps: 0
```

US-10-062-624-40 (1-293) x US-10-059-964-43 (1-882)

```
QY 1 MetAenAenLysLeuLysPheThrIleAenThrValLeuValCysLeuLeuSerLeu 20
Db 1 ATGATAATAAACTCAATTTACTATAATAAACACAGTATTAGTAGTCTTATTGTCAATTA 60
QY 21 ProAenIleSerSerLysValAlaIleAenAenAenAlaLysLysTyrTyrGlyLeuTyr 40
Db 61 CCTAATATATCTCTCAAGAGCCATAACAAATTAACGCTAAAGTAAGTACGAGTATATAT 120
QY 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAenPheSerValLysGluThr 60
Db 121 ATCAGTGACCAATATAAACCCAGTGTCTCTGTTTTCAGTAATTTTCAGTTAAAGAAC 180
QY 61 AenValIleThrLysAenLeuIleAlaLysLysLysAenValSerValSerIleGluThrLys 80
Db 181 AATGTCAATACTAAACCCCTTATAGCTTTTAAAGAGATGTTGACTCTATTGAAACCAAG 240
QY 81 ThrAenAlaSerValGlyLysSerAenProSerAenPheThrIleProTyrThrAlaVal 100
Db 241 ACTGATGCCAGTGTAGGTATTAGTAACCCATCAAAATTTTACTATCCCTATACAGCTGTA 300
QY 101 PheGlnAenAenSerValAenPheAenGlyThrIleGlyTyrThrPheAlaGluGlyThr 120
Db 301 TTTCAAGATAATCTGTCAATTTCAATGGAATCTTGGTTACACCTTTGCTGAAGGTACA 360
QY 121 ArgValGluIleGluGlySerTyrGluGluPheAenValLysAenProGlyGlyTyrThr 140
Db 361 AGAGTTGAATAGAAAGTCTTATGAGGAATTTGATGTTTAAAGAACCCCTGGAGGCTATACA 420
QY 141 LeuSerAenAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAenSerPheThr 160
Db 421 CTAAAGTGATGCCATCTGCTATTTGCAATAGCAGCTGAATGAAAGGTAATAGTTTACA 480
QY 161 ProLysGluLysValSerAenSerIlePheHisThrValMetArgAenAenProGlyLeuSer 180
Db 481 CCTAAGAAAGAAAGTTTCTAATAGTTTCTTCACTGTAATCAGAAATGATGATTAATCT 540
QY 181 IleIleSerValIleValAenValCysTyrAenPheSerLeuAenAenSerIleSer 200
Db 541 ATAATATCTGTTATAGTAATCTTGTGCTACGATTTCTCTTTGAACAAATTTGTCAATATCG 600
QY 201 ProTyrIleCysGlyValAlaGlyValAenAlaIleGluPheAenValLeuHisIle 220
Db 601 CCTACATATGTGGAGGAGCGGTAGTGTATAGTAATCTTCAATGATTTATACACATT 660
QY 221 LysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAenIleSerLeu 240
Db 661 AAGTTTGCATATCAAAAGCAAGCTAGTATTCTTCTTCACTTACCTTCACTTAAATATT 720
QY 241 PheAlaSerLeuTyrTyrHisLysValMetGlyAenGlnPheLysAenLeuAenValGln 260
```

Db 721 TTTCGTAGTTTATATACCATAAAGTAATGGCAATCAATTTAAATTTAAATGTCCAA 780  
 Qy 261 HisValAlaGluLeuAlaSerIleProlysisleThrSerAlaValAlaThrLeuAsnIle 280  
 Db 781 GATGTTGCTGAACCTTGCAAGTATACCTAAATATACATCCGAGTGTCTACACTTAATATT 840  
 Qy 281 GlyTyrPheGlyGlyGluIleGlyAlaAlaArgLeuThrPhe 293  
 Db 841 GGTATTATTTGGAGGTGAATTTGGTGCAAGATTGACATTT 879

RESULT 6  
 US-10-314-639-17  
 ; Sequence 17, Application US/10314639  
 ; Publication No. US20030103991A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rikihisa, Yasuko  
 ; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
 ; TITLE OF INVENTION: Chaffeensis  
 ; FILE REFERENCE: 22727/04021  
 ; CURRENT APPLICATION NUMBER: US/10/314.639  
 ; PRIOR FILING DATE: 2002-12-09  
 ; PRIOR APPLICATION NUMBER: US/09/314.701  
 ; PRIOR FILING DATE: 1999-05-19  
 ; NUMBER OF SEQ ID NOS: 66  
 ; SOFTWARE: Patent in ver. 2.0  
 ; SEQ ID NO 17  
 ; LENGTH: 876  
 ; TYPE: DNA  
 ; ORGANISM: Ehrlichia chaffeensis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(876)  
 US-10-314-639-17

Alignment Scores:  
 Pred. No.: 5.32e-115 Length: 876  
 Score: 1072.00 Matches: 206  
 Percent Similarity: 81.57% Conservative: 33  
 Best Local Similarity: 70.31% Mismatches: 52  
 Query Match: 71.66% Indels: 2  
 Gaps: 9

US-10-062-624-40 (1-293) x US-10-314-639-17 (1-876)

Qy 1 MetAsnLysLeuLysPheThrIleAsnThrValLeuValCysLeuSerLeu 20  
 Db 1 ATGAATAAAAAACAAGTTT--ATTATAGTACAGCATTTGGTATATTACTGTCATTA 57  
 Qy 21 ProAsnLysSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40  
 Db 58 CCTAGTGATCGTTTTCAGAGGTTACAAACAGCAGTATTAAAAACAACACTCTGGGTTATAT 117  
 Qy 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60  
 Db 118 ATTATAGTACAAATACAAACCAAGTGTCTGTTTTTATAGTAGTTTCTCAATTAAGAAGAACT 177  
 Qy 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80  
 Db 178 AACACTATACAAAAATCTTATAGGTTAAAAAAGATATTAATCTCTCTGAAGTTAAC 237  
 Qy 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaVal 100  
 Db 238 GCCGATGCTAGTCAAGGTATTAGTATCATCCAGGAATTTTACTATACCTTATATAGCAGCA 297  
 Qy 101 PheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThr 120  
 Db 298 TTGAAGATAATGCTTTTAATTTCAACGGTCTATTGGTTAC---ATTACTGAAGGCTA 354  
 Qy 121 ArgValGluIleGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThr 140  
 Db 355 AGGATTGAAATAGAGGTTCTCTATGAAGATTTTGATGCTAAAAAACCCCTGGAGGTTATGGT 414

Qy 141 LeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThr 160  
 Db 415 CTTAAATGATGCTTTTCGGTACTTTTGTCTTAGCAGCTGATATGGAAGCAACAAGTTCCAA 474  
 Qy 161 ProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSer 180  
 Db 475 CCAAAACCAAAAGCTCACAAAAGTATTTCACACTGTAAATGAAGAGTGATGGGTATCT 534  
 Qy 181 IleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSer 200  
 Db 535 ATAATATCTATCGTTTAAACGGCTGTTATGATTTTCTTCGGATAATTTATTAGTATCA 594  
 Qy 201 ProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIle 220  
 Db 595 CCTTATATATGCGAGGTATAGGTGCGATGCAATAGAAATTTTTCGCGCATTCACATT 654  
 Qy 221 LysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeu 240  
 Db 655 AAATTCGGTCCAAAGCAATTAGGCATCCTTATCAATTTATCTTATAATATCAGCTTA 714  
 Qy 241 PheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGln 260  
 Db 715 TTTGCTGATGATATTTATCATCAAGTAATAGTAACCAATTCAGAAATTTAAACGTTCAA 774  
 Qy 261 HisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIle 280  
 Db 775 CATGTAGCTGAATTAATGATGACACCTTAAAGTTTACATCTGCAGTTGCCACACTTAATGTT 834  
 Qy 281 GlyTyrPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293  
 Db 835 GGATATTTTCGGCGCTGAAGTTGGAGTAAGATTTATATTT 873

RESULT 7  
 US-10-059-964-17  
 ; Sequence 17, Application US/10059964  
 ; Patent No. US20020120115A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rikihisa, Yasuko  
 ; APPLICANT: Ohasi, No. US20020120115A1  
 ; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
 ; TITLE OF INVENTION: Chaffeensis  
 ; FILE REFERENCE: 22727/04021  
 ; CURRENT APPLICATION NUMBER: US/10/059.964  
 ; CURRENT FILING DATE: 2002-01-28  
 ; EARLIER FILING DATE: 1999-05-19  
 ; NUMBER OF SEQ ID NOS: 66  
 ; SOFTWARE: Patent in ver. 2.0  
 ; SEQ ID NO 17  
 ; LENGTH: 876  
 ; TYPE: DNA  
 ; ORGANISM: Ehrlichia chaffeensis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(876)  
 US-10-059-964-17

Alignment Scores:  
 Pred. No.: 5.32e-115 Length: 876  
 Score: 1072.00 Matches: 206  
 Percent Similarity: 81.57% Conservative: 33  
 Best Local Similarity: 70.31% Mismatches: 52  
 Query Match: 71.66% Indels: 2  
 Gaps: 12

US-10-062-624-40 (1-293) x US-10-059-964-17 (1-876)

Qy 1 MetAsnLysLeuLysPheThrIleAsnThrValLeuValCysLeuSerLeu 20  
 Db 1 ATGAATAAAAAACAAGTTT--ATTATAGTACAGCATTTGGTATATTACTGTCATTA 57  
 Qy 21 ProAsnLysSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40  
 Db 1 ATGAATAAAAAACAAGTTT--ATTATAGTACAGCATTTGGTATATTACTGTCATTA 57

```

Db 58 CCTAGTGTATCGTTTTCAGAGGTTACAAACACAGTATTAAAAACACTCTGGGTATAT 117
Qy 41 lIeSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60
Db 118 ATTAGTGACAAATACAAACCAAGTGTTCCTGTTTCTAGTAGTTTCTCAATTAAGAAACT 177
Qy 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
Db 178 AACACTATACAAAAAATCTTATACGTTTAAAAAAGATATTAACTCTCTTGAAGTTAAC 237
Qy 81 ThrAspAlaSerValGlyLysSerAsnProSerAsnPheThrIleProTyrThrAlaVal 100
Db 238 GCCGATGCTAGTCAAGGATTTAGTCAATCCAGAAATTTTACATATACCTTATATAGACGA 297
Qy 101 PheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThr 120
Db 298 TTTGAAGATAATGCTTTTAAATTTCAACGGTGTCTATTGGTTAC--ATTACTCAAGGTCTA 354
Qy 121 ArgValGluIleGlySerTyrGluGluPheAspValLysAsnProGlyTyrThr 140
Db 355 AGGATTGAATAGAAAGGTTCTTATGAAGAAATTTGATGCTTAAACCCCTGGAGGTTATGGT 414
Qy 141 LeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThr 160
Db 415 CTAAATGATGCTTTCGGTACCTTTGCTTTAGCAGTGATATGGAAGCAACAAAGTTCCNA 474
Qy 161 ProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSer 180
Db 475 CCAAAGACCAAAAGCTCAAAAAAGTATTTCACACTGTATTAAGAGAGTGTGGTTATCT 534
Qy 181 lIeIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSer 200
Db 535 APTAATCTCATCATGTTAACGGCTGTATGATATTTCTTCGGATAATTTATTAGTATCA 594
Qy 201 ProTyrIleCysGlyValAlaGlyValAspAlaIleGluPheAspValLeuHisIle 220
Db 595 CCTTATATATGTGGAGGTATAGTGTGGATGCAATAGAAATTTTTCACCATTTACATT 654
Qy 221 LysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeu 240
Db 655 AAATGCTGCTGCAAGCAAAATAGGCATCATCTTATCAATTTATATATATCATGCTTA 714
Qy 241 PheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGln 260
Db 715 TTTGCTGATGGATATTATCATCAAGTAATAGGTAACCAATTGAGAAATTTAAACGTTCAA 774
Qy 261 HisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIle 280
Db 775 CATGTAGCTGAATTAATGATGCACCTAAAGTTTACATCTGCAGTTGCCACATTAATGTT 834
Qy 281 GlyTyrPheGlyGlyLysLeuGlyAlaArgLeuThrPhe 293
Db 835 GGATATTTCGGCGCTGAAGTGTGAGTAAGATTATATTT 873

```

## RESULT 8

```

US-10-314-639-13
; Sequence 13, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikibisa, Yasuko
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/314,701
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 894
; TYPE: DNA

```

ORGANISM: Ehrlichia chaffeensis

FEATURE:

NAME/KEY: CDS

LOCATION: (1) .. (894)

US-10-314-639-13

## Alignment Scores:

Pred. No.: 1,23e-109 Length: 894  
 Score: 1026.00 Matches: 194  
 Percent Similarity: 79.59% Conservative: 40  
 Best Local Similarity: 65.99% Mismatches: 58  
 Query Match: 68.58% Indels: 2  
 Ds: 9 Gaps: 2

US-10-062-624-40 (1-293) x US-10-314-639-13 (1-894)

```

Qy 1 MetAsnAsnLysLeuLysPheThrIleIleAsnThrValLeuValCysLeuSerLeu 20
Db 13 ATGAATAGAAAAACAATTTCTTACATAATAGTACAGCAATGGTATGCTTATTGTTATA 72
Qy 21 ProAsnIleSerSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40
Db 73 CCTGTATATCATTTTCAGAACTATAAACACAGTGTCTAAAAAACACAGCTGGGTATAT 132
Qy 41 lIeSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60
Db 133 ATCAGTGGCAGTACAAACCTAGTGTTCAGTTTTTAGTAATTTTTCAGTAAAGAAACT 192
Qy 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
Db 193 AATGTTCCCAACAGCAGTAAATAGCACTTAAAGAACATTAATTCGTCAGTGGT 252
Qy 81 ThrAspAlaSerValGlyLysSerAsnProSerAsnPheThrIleProTyrThrAlaVal 100
Db 253 AGTAATGCTACTACAGGTATTAGCAATCCAGTAAATTTTACAAATTCCTTATCTACAGAA 312
Qy 101 PheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThr 120
Db 313 TTTCAAGATAATGTTGCCAATTTCAATGGGCTGTGTGTACTCTTTTCTCGATAGTCTA 372
Qy 121 ArgValGluIleGlyLysTyrGluGluPheAspValLysAsnProGlyGlyTyrThr 140
Db 373 AGAATTTGAATAGAGGATTTTCATGAAAAATTTGATGTCAAAACCCCTGGAGGTTACACA 432
Qy 141 ---LeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPhe 159
Db 433 CAAAGTAAAGATGCGTACCGTTATTTCACACTAGCACGCTGATTTAAAAAGATGGCTTCTTT 492
Qy 160 ThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeu 179
Db 493 GAACCTAAAGCGGAA---GATACAGGTGTTTATCATCTGTTATGAAAAATGATGATTA 549
Qy 180 SerIleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIle 199
Db 550 TCTATTATCTACTATGTTAAACGTCGTACGATTTTCTGTAGTAGTAATTTACAGTC 609
Qy 200 SerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHis 219
Db 610 TTACCTTATATGTGCAGGTATGATATAACGCCATAGAAATTTCTCGACGCTTTTACAT 669
Qy 220 lIeLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSer 239
Db 670 GTAAATTTGCTTACCAAGGCAAACTAGTATTAGCTATCAACTATTTACTAAAGTAAT 729
Qy 240 LeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnVal 259
Db 730 TTATTCCTTGATGGGTATTACCATCAAGTAATAGGCAATCAATTCAAAAACCTTAAACGTA 789
Qy 260 GlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsn 279
Db 790 AACCATGTTTACACACTTAAAGAAATCTCTTAAAGTCACATCTGCAGTAGTACACTTGAC 849
Qy 280 lIeGlyTyrPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293

```





```

; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; PRIORITY FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/314,701
; FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 897
; TYPE: DNA
; ORGANISM: OMP-1H
; NAME/KEY: CDS
; LOCATION: (1)..(897)
US-10-314-639-51

Alignment Scores:
Pred. No.: 2,3e-107 Length: 897
Score: 1006.50 Matches: 193
Percent Similarity: 78.72% Conservative: 40
Best Local Similarity: 65.20% Mismatches: 56
Query Match: 67.28% Indels: 7
DB: 3 Gaps: 3

US-10-062-624-40 (1-293) x US-10-314-639-51 (1-897)
Qy 1 MetAsnLysLeuLysPheThrIleAenThrValLeuValCysLeuLeuSerLeu 20
Db 1 ATGAATCACAAGAGTATGCTCTTTACATAGGTACAGCTTTGATATCCTTATGTCATTA 60
Qy 21 ProAsnLysSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40
Db 61 CCTAATGTATCATTTCTCAGGAATCATAAATACCAATGCTAACAAAT---TTAGGTATATAC 117
Qy 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60
Db 118 ATTAGTGGCAATATATAACCCAGTGTCTCTTTAGCAATTTCTCAGTAAAGAACT 177
Qy 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
Db 178 AACTTCACTACACACAGTTAGTACACTTAAAAAGATATTCATTCTGTTGACATTAGT 237
Qy 81 ThrAspAlaSerValGlyLysSerProSerAsnPheThrIleProTyrThrAlaVal 100
Db 238 ACCAATGCTGATAGCGGTATTAAATCCGAGAAATTTCACTATCCCTTATATACCAAAA 297
Qy 101 PheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThr 120
Db 298 TTTCAGACAAATGCTGCTAGTTTATGAGACACTTGGATCTTCTACGCTAGAGTTTA 357
Qy 121 ArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThr 140
Db 358 AGACTTGAATGGAAGGTTCTCTATGAAGATTTTATGATGTTAAACCCCTGGAGGATATACA 417
Qy 141 ---LeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPhe 159
Db 418 AAAGTAAAGATGATATCGTTACTTTGCCCTGGCAGCTGAGATGCAATCTGGTCAAACT 477
Qy 160 ThrProLysGluLysValSer-----AsnSerIlePheHisThrValMet 174
Db 478 TGCCCTAAACCAAGAAACATCAGGTATTCACCTCAGGTATTTATCAGCTGTTATG 537
Qy 175 ArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysTyrAspPheSerLeu 194
Db 538 AGGAATGATGGGTATCTATTCTCTCATATCAATCAATGTTGTTATTAACCTTACITTA 597
Qy 195 AsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhe 214
Db 195 AsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhe 214

```

---

```

Db 598 AGTAATCTACCAATATATCACCTTACATGTGTAGGTAGGAATAGATGCTATATACAATTT 657
Qy 215 PheAspValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeu 234
Db 658 TTTGATTCACATATATAGTTTGCACATCAAGTAAGTTAGGTATATCTATCCACTA 717
Qy 235 ProSerAsnLysSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPhe 254
Db 718 TCTTCAAAATGTTCAATTTATTTGCTGATAGCTATTTATCATAAAGTAAGTAAGTAATAAATTT 777
Qy 255 LysAsnLeuAsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAla 274
Db 778 AAAAATCTAAGGGTTCAACACGTTTATGAATTAACAACAGGTACCTAAAGTTACATCTGCT 837
Qy 275 ValAlaThrLeuAsnIleGlyTyrPheGlyGlyGluIleGlyAlaArg 290
Db 838 GTTGCTACACTTGATATTTGGGTATTTTGGTGGTGAAGTTGGAGTAAGG 885

RESULT 13
US-10-059-964-51
; Sequence 51, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; CURRENT FILING DATE: 2002-01-28
; EARLIER APPLICATION NUMBER: 09/314,701
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 897
; TYPE: DNA
; ORGANISM: OMP-1H
; NAME/KEY: CDS
; LOCATION: (1)..(897)
US-10-059-964-51

Alignment Scores:
Pred. No.: 2,3e-107 Length: 897
Score: 1006.50 Matches: 193
Percent Similarity: 78.72% Conservative: 40
Best Local Similarity: 65.20% Mismatches: 56
Query Match: 67.28% Indels: 7
DB: 3 Gaps: 3

US-10-062-624-40 (1-293) x US-10-059-964-51 (1-897)
Qy 1 MetAsnLysLeuLysPheThrIleAenThrValLeuValCysLeuLeuSerLeu 20
Db 1 ATGAATCACAAGAGTATGCTCTTTACATAGGTACAGCTTTGATATCCTTATGTCATTA 60
Qy 21 ProAsnLysSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40
Db 61 CCTAATGTATCATTTCTCAGGAATCATAAATACCAATGCTAACAAAT---TTAGGTATATAC 117
Qy 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60
Db 118 ATTAGTGGCAATATATAACCCAGTGTCTCTTTAGCAATTTCTCAGTAAAGAACT 177
Qy 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
Db 178 AACTTCACTACACACAGTTAGTACACTTAAAAAGATATTCATTCTGTTGACATTAGT 237
Qy 81 ThrAspAlaSerValGlyLysSerProSerAsnPheThrIleProTyrThrAlaVal 100
Db 238 ACCAATGCTGATAGCGGTATTAAATCCGAGAAATTTCACTATCCCTTATATACCAAAA 297
Qy 101 PheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThr 120
Db 298 TTTCAGACAAATGCTGCTAGTTTATGAGACACTTGGATCTTCTACGCTAGAGTTTA 357
Qy 121 ArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThr 140
Db 358 AGACTTGAATGGAAGGTTCTCTATGAAGATTTTATGATGTTAAACCCCTGGAGGATATACA 417
Qy 141 ---LeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPhe 159
Db 418 AAAGTAAAGATGATATCGTTACTTTGCCCTGGCAGCTGAGATGCAATCTGGTCAAACT 477
Qy 160 ThrProLysGluLysValSer-----AsnSerIlePheHisThrValMet 174
Db 478 TGCCCTAAACCAAGAAACATCAGGTATTCACCTCAGGTATTTATCAGCTGTTATG 537
Qy 175 ArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysTyrAspPheSerLeu 194
Db 538 AGGAATGATGGGTATCTATTCTCTCATATCAATCAATGTTGTTATTAACCTTACITTA 597
Qy 195 AsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhe 214
Db 195 AsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhe 214

```

```

QY 101 PheGlnAspAsnSerValAsnPhenGlyThrIleGlyTyrThrPheAlaGluGlyThr 120
Db 298 TTTCAGACAAATGCTGCTAGTTTATGTGAGCACTGGGATCTTCTACGCTAGAGGTTTA 357
QY 121 ArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThr 140
Db 358 AGACTTGAAATGGAAGGTTCTTATGAGAAATTTGATGTTAAACCCCTGGAGGATATACA 417
QY 141 ---LeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPhe 159
Db 418 AAGTAAAGAGATGCATATCGTTACTTTGGCCCTGGCAGGTGAGATGCAATCTCGTCAAACT 477
QY 160 ThrProLysGluLysValSer-----AsnSerIlePheHisThrValMet 174
Db 478 TCCTCTAAACACAAAGAAACATCAGGTATTCACCTCAGCGTATTATCAGCACTGTTATG 537
QY 175 ArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysTyrAspPheSerLeu 194
Db 538 AGGAATGATGGGTATCTATTTCACTCTGTCATATCAATGTTGTTATTAACCTTTACTTTA 597
QY 195 AsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhe 214
Db 598 AGTAATCTACCAATATCACCTTACATGTGTAGGTATGGGAATAGATGCTATACAAATTT 657
QY 215 PheAspValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeu 234
Db 658 TTGATTTCACTACATATTAAAGTTTGACATCAAGTAAGTATAGGTATTACTTACCACATA 717
QY 235 ProSerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPhe 254
Db 718 TCTTCAATGTTCAATTTATTTGCTGATAGCTATTATCATAAAGTAATAGTAAATTAATTT 777
QY 255 LysAsnLeuAsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAla 274
Db 778 AAAAATCTAAGGTTTCAACACGTTTATGAATTAACAACAGGTACCTAAAGTTACATCTGCT 837
QY 275 ValAlaThrLeuAsnIleGlyTyrPheGlyGlyGluIleGlyAlaArg 290
Db 838 GTTGCTACATCTGATATTGGGTATTTTGGTGGTGAAGTTGGAGTAAGG 885

RESULT 14
US-10-314-639-49
; Sequence 49, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/314,701
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 903
; TYPE: DNA
; ORGANISM: OMP-12
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(903)
US-10-314-639-49

```

```

Alignment Scores:
Pred. No.: 6,43e-105 Length: 903
Score: 985.50 Matches: 199
Percent Similarity: 77.41% Conservative: 34
Best Local Similarity: 66.11% Mismatches: 60
Query Match: 65.88% Indels: 9
DB: Gaps: 2

```

```

US-10-062-624-40 (1-293) x US-10-314-639-49 (1-903)
QY 1 MetAsnAsnLysLeuLysPheThrIleIleAsnThrValLeuValCysLeuLeuSerLeu 20
Db 1 ATGAAGAGAGAAAAATCAATTTATCACAATAAGTACAATATTAGTATGTTTATTGTCAATTA 60
QY 21 ProAsnIleSerSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40
Db 61 TCTAATGATCATCTTTCAACACTACAATAAGCAGCACTAAAAAACAGATTTGGGTTATAT 120
QY 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60
Db 121 GTTAGTGACATACAGCCTAGTGTCTATTATTTTACCAATTTCTCAGTAAGGAAACT 180
QY 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
Db 181 AATTTTCTCAAAAGTATCTAGCAGCTCTTAAAAAGACATTAATTTCTGTCGAATTTGAC 240
QY 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaVal 100
Db 241 GATAGTGTACTGCTGGCATTAGTTTACCCACTTAATTTTCACTACTCTCTTATAGCTGTA 300
QY 101 PheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThr 120
Db 301 TTTCAAGATATAATTTCTTAATTTTAATGGCGCTATTGGGTACACTTTTGTGAAGGCCCA 360
QY 121 ArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThr 140
Db 361 AGAATTTGAAATAGAAAGGTTCTTATGAAGAAATTCGATGTCAAAGA-CCTGGGAAGATATACA 419
QY 141 ---LeuSerAspAlaTyrArg-TyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPh 159
Db 420 GAAATCAAGATGATACCTGTTGACTTTTGGCTTTAGCAGTGATATAGACTCTATTCTCTAC 479
QY 159 eThrProLysGluLysValSerAsn-----SerIlePheHisThrVa 173
Db 480 TAGCCCAAAAATAGAACTTCACATGATGGCAACAGTTTCATATAAGGTATACCACACTGT 539
QY 173 lMetArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysTyrAspPheSe 193
Db 540 AATGAAAAATGAAGGACTATCTATAATATCCATTATGCTCAATGGCTCTCTATGATTTTTC 599
QY 193 rLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIleG1 213
Db 600 TTCAGATAATTTATCAATATTACCTTAATGATGCTGCTAGTATAGTGAATGCTATAGA 659
QY 213 uPhePheAspValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSe 233
Db 660 GTTTTTCGATGCATTACATGTTAAATTCGCGTGTCCAGGTAATATTAGGTATTACTTATCC 719
QY 233 rLeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnG1 253
Db 720 ATTATCTTCCAAACGTTAGTTTATTGCTGGTGGATATTATCACCAGTAATATGGGCAACCA 779
QY 253 nPheLysAsnLeuAsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSe 273
Db 780 ATTTAAAAATCTAAATGTTCAACATGATAGCTGAACCTTAATGACGCCACCAAGATTACATC 839
QY 273 rAlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGluIleGlyAlaArgLeuThrPh 293
Db 840 TGCAGTAGCTACACTTGACATTGGGTATTTTGGTGGTGAATTTGGAGCAAGGCTTATATT 899
QY 293 e 293
Db 900 T 900

RESULT 15
US-10-059-964-49
; Sequence 49, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20020120115A1io

```





QY 100 ValPheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyThrPheAlaGluGly 119  
 Db 298 GAATTTTCAAGACAATGCTTCAACTTCAGTGGAGCTATTGGTTATTCACCTTTTGAACAA 357  
 QY 120 ThrArgValGluIleGlySerTyrGluGluPheAspValIleAsnProGlyGlyTyr 139  
 Db 358 CTAAACATTGAAGTTGAAGTTCTTATGAAGAATTGCGATGCCAAAATCTGCTGGTTAT 417  
 QY 140 ThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMet----- 154  
 Db 418 ATTTAAATGATGATTCCTCGCTATTATTCATTGGCAGCTGAATGGACAAAGAAAAAT 477  
 QY 155 LysGlyAsnSerPheThrProLysGluLysValSer---AsnSerIlePheHisThrVal 173  
 Db 478 GATAATAAGCATCTTAGTCTTAAGGAGGAGCATGATATAAGTAAACACATATTACACAGTC 537  
 QY 174 MetArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysTyrAspPheSer 193  
 Db 538 ATGAGAAATGAAGTTATCTATATCTATATCTATATGATAAATGGCTGCTATTAATCTACCT 597  
 QY 194 LeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGlu 213  
 Db 598 CTCAATGATTATCAATATCACTTATTTTGTACAGGAATAGGTGTAGATGCTATAGAA 657  
 QY 214 PhePheAspValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSer 233  
 Db 658 TTTTGTGATGCACTGCATCTTAACTTGTCTTTCGCAAGTAAATAGGAGCTACTTACCNA 717  
 QY 234 LeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGln 253  
 Db 718 TTATCAGACACATTAGTTATTTTACAAATGGATATACCATCAAGTAAATGGAGTACNA 777  
 QY 254 PheLysAsnLeuAsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSer 273  
 Db 778 TTTTAAAACTTAAAGTCCATATATATAGTGAACCTTAAAGAGAACCCGAAATTTACATCT 837  
 QY 274 AlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGlyIleGlyAlaArgLeuThr 292  
 Db 838 GCAGTTGCTACTCTCAATGCTGATACTTTGGAGGTGAATTTGGAGTAAAGACTCACA 894

RESULT 17  
 US-10-059-964-45  
 ; Sequence 45, Application US/10059964  
 ; Patent No. US20020120115A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rikihisa, Yasuko  
 ; APPLICANT: Ohashi, No. US20020120115A1  
 ; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
 ; TITLE OF INVENTION: Chaffensis  
 ; FILE REFERENCE: 22727/04021  
 ; CURRENT APPLICATION NUMBER: US/10/059,964  
 ; CURRENT FILING DATE: 2002-01-28  
 ; EARLIER APPLICATION NUMBER: 03/314,701  
 ; EARLIER FILING DATE: 1999-05-19  
 ; NUMBER OF SEQ ID NOS: 66  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 45  
 ; LENGTH: 900  
 ; TYPE: DNA  
 ; ORGANISM: Ehrlichia canis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(900)  
 US-10-059-964-45

Alignment Scores:  
 Pred. No.: 4,548-97 Length: 900  
 Score: 918.00 Matches: 181  
 Percent Similarity: 74.25% Conservative: 41  
 Best Local Similarity: 60.54% Mismatches: 69  
 Query Match: 61.36% Indels: 8  
 DB: 12 Gaps: 4

US-10-062-624-40 (1-293) x US-10-059-964-45 (1-900)  
 QY 1 MetAsnLysLeuLysPheThrIleIleAsnThrValLeuValCysLeuLeuSerLeu 20  
 Db 1 ATGAATAGCAAGAGTAAGTTCTTTTACAATATGTACATCGTTAATATGTTTATTCATCA 60  
 QY 21 ProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40  
 Db 61 CCTAACACATCTCTCAAACTTCATAGGCAATAGTACA---AAACATCTTGGATTATAT 117  
 QY 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60  
 Db 118 GTTAGCGACATTATAGCCCGCTTCCATTTTACCAATTTTTCAGTAAAGAAACA 177  
 QY 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80  
 Db 178 AATACACATACAGTACAGTTAGTCTTAAAAAAGATGTTAATCTTATTTCTATGAAC 237  
 QY 81 ---ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAla 99  
 Db 238 ATCAGTAATGTGCTACAGGCATTAGCAAAACCAAAATTTAATCTTCTTATGTTGCA 297  
 QY 100 ValPheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGly 119  
 Db 298 GAATTTCAAGACATGCTTCACTTCACTGAGCTATTGGTTATTCATCTTTTGAACAA 357  
 QY 120 ThrArgValGluIleGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyr 139  
 Db 358 CTAAACATTGAAGTTGAAGTTCTTATGAAGAATTCGATGTCAAAATCTCTGCTGGTTAT 417  
 QY 140 ThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMet----- 154  
 Db 418 ATTTAAATGATGATTCCTCGCTATTATTCATTGGCAGCTGAATGGACAAAGAAAAAT 477  
 QY 155 LysGlyAsnSerPheThrProLysGluLysValSer---AsnSerIlePheHisThrVal 173  
 Db 478 GATAATAAGCATCTTAGTCTTAAAGGAGGAGCATGATATAAGTAAACACATATTACACAGTC 537  
 QY 174 MetArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysTyrAspPheSer 193  
 Db 538 ATGAGAAATGAAGTTATCTATATCTATATGATAAATGGCTGCTATTAATCTACCT 597  
 QY 194 LeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGlu 213  
 Db 598 CTCAATGATTATCAATATCACTTATTTTGTACAGGAATAGGTGTAGATGCTATAGAA 657  
 QY 214 PhePheAspValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSer 233  
 Db 658 TTTTGTGATGCACTGCATCTTAACTTGTCTTTCGCAAGTAAATAGGAGCTACTTACCNA 717  
 QY 234 LeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGln 253  
 Db 718 TTATCAGACACATTAGTTATTTTACAAATGGATATACCATCAAGTAAATGGAGTACNA 777  
 QY 254 PheLysAsnLeuAsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSer 273  
 Db 778 TTTTAAAACTTAAAGTCCATATATATAGTGAACCTTAAAGAGAACCCGAAATTTACATCT 837  
 QY 274 AlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGlyIleGlyAlaArgLeuThr 292  
 Db 838 GCAGTTGCTACTCTCAATGCTGATACTTTGGAGGTGAATTTGGAGTAAAGACTCACA 894

RESULT 18  
 US-10-314-639-59  
 ; Sequence 59, Application US/10314639  
 ; Publication No. US20030103991A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rikihisa, Yasuko  
 ; APPLICANT: Ohashi, No. US20030103991A1  
 ; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
 ; TITLE OF INVENTION: Chaffensis  
 ; FILE REFERENCE: 22727/04021



Db 343 GAGCTAGAGGCTCTTATGAAGAGTTCGATGTTACAGACCCCTGGAGATTGTATATAATAA 402  
Qy 143 AepLatyrArgTyPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLys 162  
Db 403 GATACCTTACAGGTACTTTGTCATTAGTAGAGAAACAAAGTGGTAAATCAT -CCCAAC 456  
Qy 163 GluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIle 182  
Db 457 GATATATGGGAA-----TATACCTGCATGAGAAATGATGGAGTATCCATTACC 504  
Qy 183 SerValIleValAsnValCysTyRaspPheSerLeuAsnAsnLeuSerIleSerProTyR 202  
Db 505 TCCGTTATATCAATGGTGTATGATCTCTTTAAAGAGCTAGAAATATACCATAT 564  
Qy 203 IleCysGlyGlyValGlyValAspAlaIleGluPheAspValLeuHisIleLysPhe 222  
Db 565 GTTTCGATTTGGTATCGGAGGAGCTTTATAGAATTTTGTAGCTTACACATTAATA 624  
Qy 223 AlaTyRGlNserLysLeuGlyIleAlaTyRserLeuProSerAsnIleSerLeuPheAla 242  
Db 625 GCATATCAAGGTAAATAGTATTAGCTATTCTTTCCACCTAGAACAAATTTATATATC 684  
Qy 243 SerLeuTyRHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGlnHisVal 262  
Db 685 GATTGTTATACATAGAGTTATAGTATCAATTTAAATTTAAATGTTCAACATGTA 744  
Qy 263 AlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIleGlyTyR 282  
Db 745 GTTGAAGCTTACAGAACACCTAAAGCTACATCTGCAATTTGCTTACACTTAATGTTAGT 804  
Qy 283 PheGlyGlyGluIleGlyAlaArgLeuThrPhe 293  
Db 805 TTCGGTGGAGAAATTGGAATTAGACTTATGTTT 837

## RESULT 20

US-10-314-639-23  
; Sequence 23, Application US/10314639  
; Publication No. US20030103991A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohagi, No. US20030103991A1io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; FILE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/314,639  
; PRIOR FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US/09/314,701  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 23  
; LENGTH: 840  
; TYPE: DNA  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(840)  
US-10-314-639-23

## Alignment Scores:

Pred. No.:	8,03e-82	Length:	840
Score:	786.50	Matches:	160
Percent Similarity:	67.92%	Conservative:	39
Best Local Similarity:	54.61%	Mismatches:	77
Query Match:	52.57%	Indels:	17
DB:	9	Gaps:	4

US-10-062-624-40 (1-293) x US-10-314-639-23 (1-840)

Qy 3 AsnLysLeuLysPheThrIleLeuAsnThrValLeuValCysLeuLeuSerLeuProAsn 22  
Db 4 AGCAAAAAAGTTTATTACAAATAGGAACAGTACTTGCATCTCTATATCACTTCTTATCT 63

Qy 23 IleSerSerLysAlaIleAsnAsnAsn-----AlaLysLysTyRtyrGlyLeuTyR 40  
Db 64 ATTGAATCCTTTTTCAGCTATATAAATCATATAACAGGAAATAACACTAGTGGTATATAT 123  
Qy 41 IleSerGlyGlnTyRAsnProSerValSerValPheSerAsnPheSerValLysGluThr 60  
Db 124 ATTACAGGGCAGTATAGACCCAGGAGTATCCCATTTTAGCAATTTTCTCAGTAAAGAAACT 183  
Qy 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80  
Db 184 AATGTTGATACATACAACTAGTAGGATATAAAAAGTGCCTCTTCTATCGATCCTTAAC 243  
Qy 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyRThrAlaVal 100  
Db 244 ACT-----TATTCAAACTTTTCAAGGTCCATATCTGTTACA 279  
Qy 101 PheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyRThrPheAlaGluGlyThr 120  
Db 280 TTTCAGATATATGCTGCTAGTTTCAGTGGAGCAATTTGGATATTCTTACCCCGAAAGTCTA 339  
Qy 121 ArgValGluIleGluGlySerTyRGlucLupheAspValLysAsnProGlyGlyTyRThr 140  
Db 340 AGACTTGAACCTGAAGGTCTTACGAAAAATTTGATGTCAAAGATCCTTAAAGACTACTCA 399  
Qy 141 LeuSerAspAlaTyRArgTyR-PheAlaLeuAlaArgGluMetLysGlyAsnSerPheThr 160  
Db 400 GCAAAAGATGCTTTTAGGTTTTTCTCTAGCAGT-----AATACGCTTACT 447  
Qy 161 ProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSer 180  
Db 448 -----ACTGTTCTCGATGCTCAAAAATATACAGTTATGAAGAATAATGGCTTATCT 498  
Qy 181 IleIleSerValIleValAsnValCysTyRaspPheSerLeuAsnAsnLeuSerIleSer 200  
Db 499 GTTGATCAATCATGATCAATGTTGTTATGATCTATCTTTTAAATAATTTAGTCGTATCA 558  
Qy 201 ProTyRileCysGlyGlyValGlyValAspAlaIleGluPheAspValLeuHisIle 220  
Db 559 CCTATATATGTCAGGATTTGGTGAAGATTTTCATTTGATTTTGTACTTTGCACATT 618  
Qy 221 LysPheAlaTyRGlNserLysLeuGlyIleAlaTyRserLeuProSerAsnIleSerLeu 240  
Db 619 AAACCTGCTTATCAAGGAAACTAGGTATTAGTTATTACTTCTTCTTAAAGATTAATGTA 678  
Qy 241 PheAlaSerLeuTyRtyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGln 260  
Db 679 TTTGCTGGTGGGTACTATCATAGAGTTATAGGGAATAAAATTTAAATAATTTAAATGTTAAC 738  
Qy 261 HisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIle 280  
Db 739 CATGTTGTACACTTGATGAATTTCTTAAAGCAACTTCTGCAGTAGCTACACTTAATGTT 798  
Qy 281 GlyTyRPhGlyGlyGluIleGlyAlaArgLeuThrPhe 293  
Db 799 GCTTATTTTGGTGGTGAAGCTGGAGTAAAGTTTACATTT 837

## RESULT 21

US-10-059-964-23  
; Sequence 23, Application US/10059964  
; Patent No. US20020120115A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohagi, No. US20020120115A1io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; FILE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/059,964  
; CURRENT FILING DATE: 2002-01-28  
; EARLIER APPLICATION NUMBER: 09/314,701  
; EARLIER FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 23

```
; LENGTH: 840
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1) ... (840)
US-10-059-964-23

Alignment Scores:
Pred. No.: 8,03e-82 Length: 840
Score: 786.50 Matches: 160
Percent Similarity: 67.92% Conservative: 39
Best Local Similarity: 54.61% Mismatches: 77
Query Match: 52.57% Indels: 17
DB: 12 Gaps: 4

US-10-062-624-40 (1-293) x US-10-059-964-23 (1-840)

QY 3 AsnLysLeuLysPheThrIleLeuValCysLeuLeuSerLeuProAsn 22
DB 4 AGCAAAAAAGTTTATTACATAGAACAGTACTTGCATCTCTATTATCATCTTATCT 63
QY 23 IleSerSerLysAlaIleAsnAsn-----AlaLysLysTyrTyrGlyLeuTyr 40
DB 64 ATTGAATCTTTTCAGCTATTAATCATATCATACAGGAATAACACTAGTGGTATATAT 123
QY 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60
DB 124 ATTACAGGCGAGTATAGACAGGAGTATCCCATTTTTCAGTAATTTCTCAGTAAGAACT 183
QY 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
DB 184 AATGTTGTATACAACTAGTAGATATAAAAAAGTGGCTCTCTATCGATCCTTAAC 243
QY 81 ThrAspAlaSerValGlyLysSerAsnProSerAsnPheThrIleProTyrThrAlaVal 100
DB 244 ACT-----TATCAAACTTTCAAGTCCATATACCTGTACA 279
QY 101 PheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThr 120
DB 280 TTTCAGATATATGCTGCTAGTTTCAGTGGAGCAATTGGATATCTTACCCCGAAGTCTA 339
QY 121 ArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThr 140
DB 340 AGACTTGAACTTGAGGTTCTTACGAAAAATTTGATGTCAAAGATCTCTAAAGACTACTCA 399
QY 141 LeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThr 160
DB 400 GCAAAAGATGCTTTTAGGTTTTTCTCTAGCAGCT-----AATACGCTACT 447
QY 161 ProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSer 180
DB 448 -----ACTGTTCTGATGCTCAAAAATATACAGTTATGAAGAAATAGGCTATCT 498
QY 181 IleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSer 200
DB 499 GTTGATCAATCATCATCAATGTTGTTATGATCTATCTTTTAAATAATTTAGTCGTATCA 558
QY 201 ProTyrIleCysGlyAlaGlyValAspAlaIleGluPheAspValLeuHisIle 220
DB 559 CCTATATATGTCAGGATTGGTGAAGATTTTCATGAATTTTTTGTACTTTGCACATT 618
QY 221 LysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeu 240
DB 619 AAACCTGCTTATCAGGAAACTAGGTATATTACTTCTTCTTAAAGTAATGTA 678
QY 241 PheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGln 260
DB 679 TTTGCTGGGTACTATCATAGAGTTATAGGAATAAATTTAAAAATTTAAATGTTAAC 738
QY 261 HisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrIleAsnIle 280
DB 739 CATGTTGTACACTTGATGAATTTCCATAAGCAACTCTCGAGTAGCTACACTTAAATGTT 798
```

```
QY 281 GlyTyrPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293
DB 799 GCTTATTGTTGGTGAAGCTGGAGTAAGATTACATTT 837

RESULT 22
US-10-314-639-29
; Sequence 29, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20030103991A1
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; PRIORITY FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/314,701
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patent In Ver. 2.0.
; SEQ ID NO 29
; LENGTH: 858
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1) ... (858)
US-10-314-639-29

Alignment Scores:
Pred. No.: 1.62e-81 Length: 858
Score: 784.00 Matches: 158
Percent Similarity: 70.07% Conservative: 48
Best Local Similarity: 53.74% Mismatches: 78
Query Match: 52.41% Indels: 10
DB: 9 Gaps: 4

US-10-062-624-40 (1-293) x US-10-314-639-29 (1-858)

QY 1 MetAsnLysLeuLysPheThrIleLeuAsn---ThrValLeuValCysLeuLeuSer 19
DB 1 ATGAATATAGAAAAAGTTTTTTTATAGTGCATCATTACTAGCAAGCTTTATTATTC 60
QY 20 LeuProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeu 39
DB 61 ACATCTGAGGCTCTTCTACAGGAAATGTAGTAACCATACTATTATTAACCTAGGTTA 120
QY 40 TyrIleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGlu 59
DB 121 TATATCAGTGGACAATATAGACAGGAGTTTCTCATTTTAGCAAAATTTTCAGTCAAGAA 180
QY 60 ThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThr 79
DB 181 ACCAACTACATACTACTCAACTAGTTGGGCTTAAAGGACATCAGTGTCTATA----- 234
QY 80 LysThrAspAlaSerValGlyLysSerAsnProSerAsnPheThrIleProTyrThrAla 99
DB 235 -----GGGACAGTAATATCAACACCTACACAAATTTCAACTTCTCTTACATGCA 285
QY 100 ValPheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGly 119
DB 286 GAAATTTCAAGACAATGCCATAAGTTTCAGTGGGCAATTTGGATATCTTGTATTCGAGAA 345
QY 120 ThrArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyr 139
DB 346 TTTAGAATTGAAGTAGAGGCTTCTTATCAAGAAATTTGATGTATAAAATTCAGAAAGGA 402
QY 140 ThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPhe 159
DB 403 TCTCTCAGACAGCATACAGGTATTTTGCATCAGCTAGCTGCTATGATGGATGGC----- 453
QY 160 ThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeu 179
```

454	Db	 ACCTAATAAACTCTAGTCCTGATGCACACAGAAAAATTCACTGTCATGAGAAATGACGGTTTA	513
180	QY	 SerIleSerValIleValAsnValCysTyrAspPheSerLeuAsnLeuSerIle	199
514	Db	 TCAATTTTCATCAGTAATGATAAATGGGTGTTCACAAATTTACATTAGATGATATACCAGTA	573
200	QY	 SerProTyrIleCysGlyGlyValaspAlaIleGluPhePheaspValLeuHis	219
574	Db	 GTACCGTATGATGCGCAGGAATAGGAGGAGATTTTCATAGAGTTTTTAAATGATTTACAT	633
220	QY	 IleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSer	239
634	Db	 GTTAGTTTCGTATCATCAGGCACAGGTAGGTATTAGTTATTCTATATCCCTCGAAGTAAGT	693
240	QY	 LeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnVal	259
694	Db	 TTATTCTTTAACGGATATTACCATAAAGTAAACAGGTAAACAGATTTTAAAACTTTACACGTT	753
260	QY	 GlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsn	279
754	Db	 CAACACGTAAAGTAAATTAAGTACGCTCCTAAGTTTCACATCTGCGAGTTGCTACACTCAAT	813
280	QY	 IleGlyTyrPheGlyGlyGluIleGlyAlaArgLeuThrPhe	293
814	Db	 GTTGGGTACTTTGGTGGCGCAATTTGGAGTAAAGTATTTATTTT	855

RESULT 23

```

US-10-059-964-29
; Sequence 29, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20020120115A1io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; CURRENT FILING DATE: 2002-01-28
; EARLIER APPLICATION NUMBER: 09/314,701
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
; LENGTH: 858
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(858)
US-10-059-964-29

```

## Alignment Scores:

Pred. No.: 1.17e-79 Length: 852  
Score: 768.00 Matches: 161  
Percent Similarity: 68.24% Conservative: 41  
Best Local Similarity: 54.39% Mismatches: 78  
Query Match: 51.34% Indels: 16  
DB: 9 Gaps: 6

US-10-062-624-40 (1-293) x US-10-314-639-25 (1-852)

Qy 1 MetAsnLeuLysLeuLysPheThrIleAlaLeuValCysLeuLeuSer--- 19  
Db 1 ATGAGTCTAAAAAAGCTTTTATAATAGGTCAGTGTAGTATGTTAGTGTCAATC 60  
Qy 20 LeuPro-----AsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyr 37  
Db 61 TTACCTACTAAATCTTTGTCACAACTTAATAATATATAACACT---AAGTGCACT 117  
Qy 38 GlyLeuTyrIleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerVal 57  
Db 118 GGGCTATATGTCAGTGGCAATATAAACCTACTGTTCTCACTTTAGTAATTTTTCAC 177  
Qy 58 LysGluThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIle 77  
Db 178 AAAGAACTTATCTGACACTAAAGAGTTATAGGACTAGCAAGAGATATTAAGTCTATT 237  
Qy 78 GluThrLysThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyr 97  
Db 238 -----ACAGAT-----ATAACAACAATAAATAATTAACACTTCTTAT 276  
Qy 98 ThrAlaValPheGlnAsnAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAla 117  
Db 277 AACACAAATTTCAAGATAATGCTGTTAGCTTCAGTCAGCTGTTGGATATATTTCCCAA 336  
Qy 118 GluGlyThrArgValGluIleGlySerTyrGluGluPheAspValLysAsnProGly 137  
Db 337 GACAGTCCAGGGTTGAGTGAATGCTTATGAAGAATTTGACGTTAAATAATTCCTGGT 396  
Qy 138 GlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsn 157  
Db 397 AATTACGTAGTAAGTGAAGCTTCAGGTATATGCTTTAGCAGAGGAATGATAAT--- 453  
Qy 158 SerPheThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAsp 177  
Db 454 -----CTTCAAAAATATCCTGAAACAATAAAGTATGTTGTTATAAAGACAAT 501  
Qy 178 GlyLeuSerIleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeu 197  
Db 502 GGCTTATCTGCGCATCATTAATCAATGGCTGTTAAGATTTTCTTTAAACAATTTA 561  
Qy 198 SerIleSerProTyrIleCysGlyAlaGlyValAspAlaIleGluPhePheAspVal 217  
Db 562 AAAGTATCACCTTACATATGCTAGGTTTGGTGGGACATATAGAATTTTGTAGTCT 621  
Qy 218 LeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsn 237  
Db 622 GTAAAGTTTAAATTTGCTTATCAAGGTAAAGTAGTATCAGTATCCATTATCTCTAAT 681  
Qy 238 IleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeu 257  
Db 682 ATGATATATTGCTGACGGATATATACCATAGGTATGAGGAATAAATTTAAACAATTTA 741  
Qy 258 AsnValGlnHisValAlaGluAlaSerIleProLysIleThrSerAlaValAlaThr 277  
Db 742 AATGTTCAACAGCTGTTAGTCTTAAACAGTCATCTCAAGTCTACTTTTGGCAGTACT 801  
Qy 278 LeuAsnIleGlyTyrPheGlyGluIleGlyAlaArgLeuThrPhe 293  
Db 802 CTTAATGTTGAGTATTTCCGTAGTGAATTTGGGTTAAATTTATATTT 849

RESULT 25

US-10-059-964-25

; Sequence 25, Application US/10059964  
; Patent No. US20020120115A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20020120115A1  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; TITLE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/059,964  
; CURRENT FILING DATE: 2002-01-28  
; EARLIER APPLICATION NUMBER: 09/314,701  
; EARLIER FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 25  
; LENGTH: 852  
; TYPE: DNA  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(852)  
US-10-059-964-25

Alignment Scores:  
Pred. No.: 1.17e-79 Length: 852  
Score: 768.00 Matches: 161  
Percent Similarity: 68.24% Conservative: 41  
Best Local Similarity: 54.39% Mismatches: 78  
Query Match: 51.34% Indels: 16  
DB: 12 Gaps: 6

US-10-062-624-40 (1-293) x US-10-059-964-25 (1-852)

Qy 1 MetAsnLeuLysLeuLysPheThrIleAlaLeuValCysLeuLeuSer--- 19  
Db 1 ATGAGTCTAAAAAAGCTTTTATAATAGGTCAGTGTAGTATGTTAGTGTCAATC 60  
Qy 20 LeuPro-----AsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyr 37  
Db 61 TTACCTACTAAATCTTTGTCACAACTTAATAATATATAACACT---AAGTGCACT 117  
Qy 38 GlyLeuTyrIleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerVal 57  
Db 118 GGGCTATATGTCAGTGGCAATATAAACCTACTGTTCTCACTTTAGTAATTTTTCAC 177  
Qy 58 LysGluThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIle 77  
Db 178 AAAGAACTTATCTGACACTAAAGAGTTATAGGACTAGCAAGAGATATTAAGTCTATT 237  
Qy 78 GluThrLysThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyr 97  
Db 238 -----ACAGAT-----ATAACAACAATAAATAATTAACACTTCTTAT 276  
Qy 98 ThrAlaValPheGlnAsnAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAla 117  
Db 277 AACACAAATTTCAAGATAATGCTGTTAGCTTCAGTCAGCTGTTGGATATATTTCCCAA 336  
Qy 118 GluGlyThrArgValGluIleGlySerTyrGluGluPheAspValLysAsnProGly 137  
Db 337 GACAGTCCAGGGTTGAGTGAATGCTTATGAAGAATTTGACGTTAAATAATTCCTGGT 396  
Qy 138 GlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsn 157  
Db 397 AATTACGTAGTAAGTGAAGCTTCAGGTATATGCTTTAGCAGAGGAATGATAAT--- 453  
Qy 158 SerPheThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAsp 177  
Db 454 -----CTTCAAAAATATCCTGAAACAATAAAGTATGTTGTTATAAAGACAAT 501  
Qy 178 GlyLeuSerIleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeu 197  
Db 502 GGCTTATCTGCGCATCATTAATCAATGGCTGTTAAGATTTTCTTTAAACAATTTA 561





```
DB: 12 Gaps: 3
US-10-062-624-40 (1-293) x US-10-059-964-57 (1-846)
QY 1 MetAsnAspLysLeuLysPheThrIleAlaAsnThrValLeuValCysLeuSerLeu 20
DB 1 ATGATATATAAAGAAATTTTATAGTATGCTCTATGATATAAATCTACTATG 60
QY 21 ProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyThrGlyLeuTy 40
DB 61 CCAATTGATGCTCTCTATGGAAGTACATAATATACACATTTTACACCTAGGCTGAT 120
QY 41 IleSerGlyGlnTyLysProSerValSerValPheSerAsnPheSerValLysGluThr 60
DB 121 ATTAGTGGCAATACAGCCAGGAGTTTCCCACTTTTACAAATTTTTCAGTCAAGAAACA 180
QY 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
DB 181 CATTGTAATAGTGCATTAATGCTGGCTACAAAGATATA-----AAA 225
QY 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyThrAlaVal 100
DB 236 GTAACATAATACAGTAGTATCAACAAATACTAGTTTAACTTTTCCCTATGTCAGAA 285
QY 101 PheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyThrPheAlaGluGlyThr 120
DB 286 TTTCAGATAACGCAATGAGCTTTAGTGGCAATAGGATGCTTTTATTCAGAGGATCTACT 345
QY 121 ArgValGluIleGluGlySerTyGluGluPheAspValLysAsnProGlyGlyTyThr 140
DB 346 AGAATTGAAGTAGAAGCTTCTATGAAGATTTGACGTTTAAATTCCTGAAGGATCTACT 405
QY 141 LeuSerAspAlaTyArgTyThrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThr 160
DB 406 ---ACAGACTCCTATAGATATTTCCGTTAGCAGTGGCATGGTGAATAATATTCCT 462
QY 161 ProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSer 180
DB 463 ACAAGTCAAAA-----TTTACTGTAATGAGAAACGACGGGTATTA 504
QY 181 IleIleSerValIleValAsnValCysTyArgPheSerLeuAsnAsnLeuSerIleSer 200
DB 505 ATCTCATCTGTTATGATAAATGCTGTTACAAATGTCATACATAATGATATACAGCAGAA 564
QY 201 ProTyIleCysGlyGlyValaGlyValAspAlaIleGluPheAspValLeuHisIle 220
DB 565 CCTTACATATGTCAGGAGTACAGGAGATTTATAGATTTCTTCAATGGCTTTCACTGTT 624
QY 221 LysPheAlaTyThrGlnSerLysLeuGlyIleAlaTyThrSerLeuProSerAsnIleSerLeu 240
DB 625 AAGCTAGCTTATCAAGGTAAAGTAGGCAATAGTTATCAATATTCCTGGAAGTAAGATTA 684
QY 241 PheAlaSerLeuTyThrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGln 260
DB 685 TTTATTGATGATCTACCAATAAGTAAAGCAAGTTTAAATAATTTACAGCTTCAA 744
QY 261 HisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIle 280
DB 745 CATGTAGTGCACTTGACAGCACTCCCTAAAGTTACATCTGCAGTTGCACACTTAATATT 804
QY 281 GlyTyPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293
DB 805 GGATACCTTTGGTTGTGAAGCTGGAGTAAGATTCATATTT 843
RESULT 28
US-10-314-639-27
; Sequence 27, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohsei, No. US20030103991A1
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
```

```
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/314,701
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patent in ver. 2.0
; SEQ ID NO 27
; LENGTH: 828
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(828)
; US-10-314-639-27
Alignment Scores:
Pred. No.: 3,93e-76 Length: 828
Score: 737.50 Matches: 161
Percent Similarity: 65.29% Conservative: 29
Best Local Similarity: 55.33% Mismatches: 84
Query Match: 49.30% Indels: 17
DB: Gaps: 6
US-10-062-624-40 (1-293) x US-10-314-639-27 (1-828)
QY 3 AsnLysLeuLysPheThrIleAlaAsnThrValLeuValCysLeuSerLeuProAsn 22
DB 4 AGTAAAAAATTTTATACATAGGAGCAACATTTATCATATGTG--TTACCTAAC 60
QY 23 IleSerSerLysAlaIleAsnAsnAlaLysLysTyThrGlyLeuTyIleSer 42
DB 61 ATATCTTTTCCAGAACTATTAACAATAACTGATAAACTTTCTGGGTTATATATAAGT 120
QY 43 GlyGlnTyLysProSerValSerValPheSerAsnPheSerValLysGluThrAsnVal 62
DB 121 GGGCAATATAAACAGGAGTTTCTCATTTTCAGCAAAATTTTCAGTCAAAAGAAATCTATAAT 180
QY 63 IleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLysThrAsp 82
DB 181 GATAACATTCACATTAATTTGGTTAAGACAC-----AACGCAATTTCTACTAGTACC--- 231
QY 83 AlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyThrAlaValPheGln 102
DB 232 -----CTTAATATTAATACAGATTTTAAATATCCCTATAAAGTAACATTTCAA 279
QY 103 AsnAsnSerValAsnPheAsnGlyThrIleGlyTyThrPheAlaGluGlyThrArgVal 122
DB 280 AATAACATTAACAGCTTTAGTGGAGCTATTGGTTTATTTCTGATCCACAGGGGCAAGATTT 339
QY 123 GluIleGluGlySerTyGluGluPheAspValLysAsnProGlyGlyTyThrLeuSer 142
DB 340 GAGCTTGAGGTTCTTATGAAGAAATTTGATGTGACAGATCTCTGGAGACTGCTTAATAAAA 399
QY 143 AspAlaTyArgTyThrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLys 162
DB 400 GATACCTATAGATATTTCCGTTTAGCTAGAAACCCATCAGGTTCTAGCCCTACC----- 453
QY 163 GluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIle 182
DB 454 -----TCAAAACAAC-----TATACCTGTTATGAGAAATGATGGTGTTCATTTACT 498
QY 183 SerValIleValAsnValCysTyArgPheSerLeuAsnAsnLeuSerIleSerProTyThr 202
DB 499 TCTGTTATATTAATGCTGTTATGACATCTTTTAAAGGATTTGAAGATTAACCTTAT 558
QY 203 IleCysGlyGlyAlaGlyValAspAlaIleGluPheAspValLeuHisIleLysPhe 222
DB 559 GTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 618
QY 223 AlaTyGlnSerLysLeuGlyIleAlaTyThrSerLeuProSerAsnIleSerLeuPheAla 242
DB 619 GCATACCAGCAAGTAGTGTATGATATATATATATATATATATATATATATATATATAT 678
```

QY 243 SerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGlnHisVal 262  
 DB 679 GATGATATATCATAGGTTATAGGAATCAATTCACAACTCTAAATGTTCAACACGCG 738  
 QY 263 AlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIleGlyTyr 282  
 DB 739 GCTAGTACAGATTTGGACCT-----GTATACGAGTAGGCACACTTAACTTACATGTTAT 792  
 QY 283 PheGlyGlyGluIleGlyAlaArgLeuThrPhe 293  
 DB 793 TTTGGTGGTGAATCGGAATTAGACTTACATTT 825

## RESULT 29

US-10-059-964-27  
 ; Sequence 27, Application US/10059964  
 ; Patent No. US20020120115A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rikihisa, Yasuko  
 ; APPLICANT: Ohasi, No. US20020120115A1  
 ; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
 ; FILE REFERENCE: 22727/04021  
 ; CURRENT APPLICATION NUMBER: US/10/059,964  
 ; CURRENT FILING DATE: 2002-01-28  
 ; EARLIER APPLICATION NUMBER: 09/314,701  
 ; EARLIER FILING DATE: 1999-05-19  
 ; NUMBER OF SEQ ID NOS: 66  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 27  
 ; LENGTH: 828  
 ; TYPE: DNA  
 ; ORGANISM: Ehrlichia chaffeensis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(828)  
 US-10-059-964-27

Alignment Scores:  
 Pred. No.: 3,93e-76 Length: 828  
 Score: 737.50 Matches: 161  
 Percent Similarity: 65.29% Conservative: 29  
 Best Local Similarity: 55.33% Mismatches: 84  
 Query Match: 49.30% Indels: 17  
 DB: 12 Gaps: 6

US-10-062-624-40 (1-293) x US-10-059-964-27 (1-828)

QY 3 AsnLysLeuLysPheThrIleLeuValLeuValCysLeuLeuSerLeuProAsn 22  
 DB 4 AGTAAAAAAATTTTATACATAGGAGCAACACTTATTCATATGTTG---TTACTTAA 60  
 QY 23 IleSerSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyrIleSer 42  
 DB 61 ATATCTTTCCAGAACTATTAACTAATACACTGATAACTTCTCGGTTATATAAGT 120  
 QY 43 GlyIleTyrLysProSerValSerValPheSerAsnPheSerValLysGluThrAsnVal 62  
 DB 121 GGGCAATATAAACCCAGGGATTTCTCATTTTCAGCAAAATTTTCAGTCAAAAGAAATCTATAAT 180  
 QY 63 IleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLysThrAsp 82  
 DB 181 GATAACATTCACATTAATGGTTAGACAC-----AACGCAATTTCTACTAGTACC--- 231  
 QY 83 AlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaValPheGln 102  
 DB 232 -----CTTAATATTAAACAGATTTTAAATATCCCTATAAAGTAACATTTCAA 279  
 QY 103 AspAsnSerValAsnPheAsnGlyThrIleGlyTyrPheAlaGluGlyThrArgVal 122  
 DB 280 AATAACATTCACAGCTTTAGTGAGCTATTGGTTATTTCTGATCCCAACGGGCAAGATTT 339  
 QY 123 GluIleGluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThrLeuSer 142

DB 340 GAGCTTGAGGTTCTTATGAAGAAATTTGATGTGCAGACTCTCGAGACTGCTGTTATAAAA 399  
 QY 143 AspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLys 162  
 DB 400 GATACCTATAGATATTTCGCTTTAGCTAGAAACCCATCAGGTTCTAGCCCTACC----- 453  
 QY 163 GluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIle 182  
 DB 454 -----TCAAAACAAC-----TATACCTTTATGAGAAATGATGGTGTTCATTA 498  
 QY 183 SerValIleValAsnValCysTyrAspPheSerLeuAsnLeuSerIleSerProTyr 202  
 DB 499 TCTGTTATTTAATGCTGTATGACATCTTTTAAAGGATTTAGAGTATACCTTAT 558  
 QY 203 IleCysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIleLysPhe 222  
 DB 559 GTATGTGTGGTGTAGGTGGAGATTTTATAGAAATTTTTCAGCGATTACACATTAATA 618  
 QY 223 AlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAla 242  
 DB 619 GCATACCAAGCAAGTTAGGTATCAATTTATCATCTATCGACTCAAGCAAGCGTATTATT 678  
 QY 243 SerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGlnHisVal 262  
 DB 679 GATGATATTTATCATAGGTTATAGGAATCAATTCACAACTCTAAATGTTCAACACG 738  
 QY 263 AlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIleGlyTyr 282  
 DB 739 GCTAGTACAGATTTTGGACCT-----GTATACGAGTAGGCACACTTAACTTAT 792  
 QY 283 PheGlyGlyGluIleGlyAlaArgLeuThrPhe 293  
 DB 793 TTTGGTGGTGAATCGGAATTAGACTTACATTT 825

## RESULT 30

US-10-314-639-53  
 ; Sequence 53, Application US/10314639  
 ; Publication No. US20030103991A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rikihisa, Yasuko  
 ; APPLICANT: Ohasi, No. US20030103991A1  
 ; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
 ; FILE REFERENCE: 22727/04021  
 ; CURRENT APPLICATION NUMBER: US/10/314,639  
 ; CURRENT FILING DATE: 2002-12-09  
 ; PRIOR APPLICATION NUMBER: US/09/314,701  
 ; PRIOR FILING DATE: 1999-05-19  
 ; NUMBER OF SEQ ID NOS: 66  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 53  
 ; LENGTH: 882  
 ; TYPE: DNA  
 ; ORGANISM: p30-6  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(882)  
 US-10-314-639-53

Alignment Scores:  
 Pred. No.: 1.9e-75 Length: 882  
 Score: 732.00 Matches: 158  
 Percent Similarity: 67.46% Conservative: 41  
 Best Local Similarity: 53.56% Mismatches: 89  
 Query Match: 48.93% Indels: 8  
 DB: 9 Gaps: 4

US-10-062-624-40 (1-293) x US-10-314-639-53 (1-882)

QY 1 MetAsnAsnLysLeuLysPheThrIleLeuValLeuValCysLeuSerLeu 20  
 DB 13 ATGTACAAAAAATACAAAATAATGACAGCGGTGTAGTATTATTTCACATGTTATTCTTA 72

```
QY 21 ProhenleSerSerSerLysAlaIleAenAenAenAlaLysLysTyTyTyGlyLeuTy 40
Db 73 CCTCATGTTCTTTCGCAAAAATACAAACAGCAATAAA-----CTTGGATTATAC 123

QY 41 IleSerGlyGlnTyTyLysProSerValSerValPheSerAenPheSerValLysGluThr 60
Db 124 ATCAGTGACAGATATAACCTAGTGTCTCTGTTTTTATAGCAATTTTTCAGCAAAAGAAACC 183

QY 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
Db 184 AATGTTTCATACAGTACAACTCATGCGCTTAAAAAGACATTTCTATTGAAAGGATTA 243

QY 81 ThrAspAlaSerValGlyIleSerAsnProSerAenPheThrIleProTyThrAlaVal 100
Db 244 ACTGGAATAGCGAGGTTTCTTATGAAAAATTTGATGCTAAAGACCTTGGTGAGTACACC 303

QY 141 ---LeuSerAspAlaTyArgTyPheAlaLeuAlaArgGluMetLysGlyAsnSerPhe 159
Db 424 AAAATAAAGATGCTTATAGATATTTGCTCTAGTACGTGAATTCATGCTCTCAATT 483

QY 160 ThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeu 179
Db 484 TATCCAAAGAGATAAT---AACACAGGAACACATTATATCTGTTATGAGAAATGATGGTATA 540

QY 180 SerIleIleSerValIleValAsnValCysTyTyAspPheSerLeuAsnAsnLeuSerIle 199
Db 541 TCTATTTCTTCTGCTACAGTAAATGGCTGCTATGATTTCTTTTCCAGTTTATCTTCTGTC 600

QY 200 SerProTyThrLysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHis 219
Db 601 ACC---TATATGTGTATAGGCATCGGTATAGATCTAGTAAATTTCTTAATGC-ATACAT 656

QY 220 IleLysPheAlaTyGlnSerLysLeuGlyIle-AlaTySerLeuProSerAsnIleSe 239
Db 657 ATTAAGTTTGTCTCCCAAGGTAGTTAAGTGTAACTTTATCTGTATCTCCCAATGTTAA 716

QY 239 rLeuPheAlaSerLeuTyTyHisLysValMetGlyAsnGlnPheLysAsnLeuAsnVa 259
Db 717 TTTATTTGCAGATGATATATCATAAAGTATGATGGCAATAAATTTAAAAATTTACCTGT 776

QY 259 lGlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAs 279
Db 837 TATTGGCTACCTCGGTGTAATTTGGCATAAGATTTTATATTT 879

RESULT 31
US-10-059-964-53
; Sequence 53, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffensis
; FILE REFERENCE: 27277/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; CURRENT FILING DATE: 2002-01-28
; EARLIER APPLICATION NUMBER: 09/314,701
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patent In Ver. 2.0
```

```
; SEQ ID NO 53
; LENGTH: 882
; TYPE: DNA
; ORGANISM: p30-6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(882)
US-10-059-964-53

Alignment Scores:
Pred No.: 1,9e-75 Length: 882
Score: 732.00 Matches: 158
Percent Similarity: 67.46% Conservative: 41
Best Local Similarity: 53.56% Mismatches: 89
Query Match: 48.93% Indels: 8
DB: 12 Gaps: 4

US-10-062-624-40 (1-293) x US-10-059-964-53 (1-882)

QY 1 MethAenLysLysLeuLysPheThrIleAenThrValLeuValCysLeuLeuSerLeu 20
Db 13 ATGTACAAAAATACAACTAATGACAGCAGGTGTAGTATTATTTCACATGTTATTCTTA 72

QY 21 ProhenleSerSerSerLysAlaIleAenAenAenAlaLysLysTyTyTyGlyLeuTy 40
Db 73 CCTCATGTTCTTTCGCAAAAATACAAACAGCAATAAA-----CTTGGATTATAC 123

QY 41 IleSerGlyGlnTyTyLysProSerValSerValPheSerAenPheSerValLysGluThr 60
Db 124 ATCAGTGACAGTATAACCTAGTGTCTCTGTTTTTATAGCAATTTTTCAGCAAAAGAAACC 183

QY 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
Db 184 AATGTTTCATACAGTACAACTCATGCGCTTAAAAAGACATTTCTATTGAAAGTTCAT 243

QY 81 ThrAspAlaSerValGlyIleSerAsnProSerAenPheThrIleProTyThrAlaVal 100
Db 244 ACTGGAATAGCGAGGTTTCTTATGAAAAATTTGATGCTAAAGACCTTGGTGAGTACACC 303

QY 101 PheGlnAspAenSerValAenPheAsnGlyThrIleGlyTyThrPheAlaGluGlyThr 120
Db 304 TTTCAAGATATATGTTGCTGCTAGCGGTGACCTTGATTTCTTTATCTTAAGGATTA 363

QY 121 ArgValGluIleGluGlySerTyTyGluGluPheAspValLysAsnProGlyTyThr 140
Db 364 AGGATTGAAATGGGGTTTTCTTATGAAAAATTTGATGCTAAAGACCTTGGTGAGTACACC 423

QY 141 ---LeuSerAspAlaTyArgTyPheAlaLeuAlaArgGluMetLysGlyAsnSerPhe 159
Db 424 AAAATAAAGATGCTTATAGATATTTGCTCTAGTACGTGAATTCATGCTCTCAATT 483

QY 160 ThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeu 179
Db 484 TATCCAAAGAGATAAT---AACACAGGAACACATTATATCTGTTATGAGAAATGATGGTATA 540

QY 180 SerIleIleSerValIleValAsnValCysTyTyAspPheSerLeuAsnAsnLeuSerIle 199
Db 541 TCTATTTCTTCTGCTACAGTAAATGGCTGCTATGATTTCTTTTCCAGTTTATCTTCTGTC 600

QY 200 SerProTyThrLysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHis 219
Db 601 ACC---TATATGTGTATAGGCATCGGTATAGATCTAGTAAATTTCTTAATGC-ATACAT 656

QY 220 IleLysPheAlaTyGlnSerLysLeuGlyIle-AlaTySerLeuProSerAsnIleSe 239
Db 657 ATTAAGTTTGTCTCCCAAGGTAGTTAAGTGTAACTTTATCTGTATCTCCCAATGTTAA 716

QY 239 rLeuPheAlaSerLeuTyTyHisLysValMetGlyAsnGlnPheLysAsnLeuAsnVa 259
Db 717 TTTATTTGCAGATGATATATCATAAAGTATGATGGCAATAAATTTAAAAATTTACCTGT 776

QY 259 lGlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAs 279
Db 837 TATTGGCTACCTCGGTGTAATTTGGCATAAGATTTTATATTT 879
```

```
Db      777 TCAATACGTTAATACTTTAGAGAGTATCCAAAGATTACATCTGCAATTGCTACACTTGA 836
Qy      nileGlyTyrPheGlyGlyGluileGlyAlaArgLeuThrPhe 293
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      837 TATTGGCTACCTCGGTGGTGAATGGCATAAGATTATATATT 879

RESULT 32
US-10-314-639-21
; Sequence 21, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20030103991A1io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; PRIOR FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/314,701
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 21
; LENGTH: 888
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
; NAME/KEY: CDS
; LOCATION: (1)..(888)
US-10-314-639-21

Alignment Scores:
Pred. No.: 5,33e-68 Length: 888
Score: 668.00 Matches: 139
Percent Similarity: 63.0% Conservative: 49
Best Local Similarity: 46.64% Mismatches: 94
Query Match: 44.65% Indels: 16
DB: 9 Gaps: 4

US-10-062-624-40 (1-293) x US-10-314-639-21 (1-888)
Qy      6 LysPheThrIleAlaSerValLeuValCysLeuSerLeuProAsnIleSerSer 25
Db      10 AAATTTAAATTTGTAATGTTATTAACATTTTGTATTTCTTTTCCCACTTAAGTCA 69
Qy      26 -----SerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyrIleSerGly 43
Db      70 TTTACAACATATGCAAAATAAACAATCACTCAAAAAGTTGGATTGTACATAAGTGT 129
Qy      44 GlnTyrLysProSerValSerValPheSerAsnPheThrAlaValPheGlnAsp 63
Db      130 CAATAAAGCCCAAGTATCTCTCATTTCAAGAATTTTTCAGTAGAAGAAATGACAAAGTA 189
Qy      64 ThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLysThrAspAla 83
Db      190 GTA---GATTGTATGCTTTCACACTGATGTTACATATATACAGAACATATA----- 240
Qy      84 SerValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaValPheGlnAsp 103
Db      241 -----TTACGAGATAATACAAATTCACACCTCATTTATTTGCAAAAGTTCAAGAAC 291
Qy      104 AsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArgValGlu 123
Db      292 AAATTTATAAATTTACAGCAGTGCATTTGTTATTTATCTCGGCAAGGACCAAGGTTAGAA 351
Qy      124 IleGluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThrLeuSerAsp 143
Db      352 ATAGAAAGCTCTTATGGGATTTTGATGTTGTAATATATAAAATATATGTCAGTACAGAT 411
Qy      144 AlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLys--- 162
Db      412 GTTAATAGATAATTTTGTCTTAGTACGTGAAGAAATGTTTCAAAATTTCTCTCCAAACCA 471
```

```
Qy      163 -----GluLysValSerAsnSerIlePheHisThrValMetArg 175
Db      472 CATGAACTAGTCAACCTCTGACAGCATTAATCTCTAAAAGTCTTTTATACATTTAATGAAG 531
Qy      176 AsnAspGlyLeuSerIleIleSerValIleValAsnValCysTyrAspPheSerLeuAsn 195
Db      532 AATAATGGGTATTGTTGTCATAGTAATAATCAACGGTTGTTATGATTTTCTTTAAT 591
Qy      196 AsnLeuSerIleSerProTyrIleCysGlyAlaGlyValAspAlaIleGluPhePhe 215
Db      592 AACACAAATATCACCTTACCTATGATAGAGTTGGAGGAGATTTTATAGAGTTTTTT 651
Qy      216 AspValLeuHisIleLeuPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuPro 235
Db      652 GAAGTAATGCATATCAAGTTTGTGTCCTGCAAGTAAAGTTGTTATGATCTATCCAAATCT 711
Qy      236 SerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLys 255
Db      712 CCTCTATTACTATTTTGTCTGATGCAVATTTATCACAAGGTCAATAATAATAATTAAC 771
Qy      256 AsnLeuAsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAlaVal 275
Db      772 AACCTACATGTTAAGTATTATCACTTAACTTAAACCTCACCTACCATTTACCTCTGCAACA 831
Qy      276 AlaThrLeuAsnIleGlyTyrPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293
Db      832 GCCAACTAAACATTTGAATATTTTGTGTTGAAGTTGGGATGAGATTTATATTT 885

RESULT 33
US-10-059-964-21
; Sequence 21, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20020120115A1io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; CURRENT FILING DATE: 2002-01-28
; EARLIER APPLICATION NUMBER: 09/314,701
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 21
; LENGTH: 888
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
; NAME/KEY: CDS
; LOCATION: (1)..(888)
US-10-059-964-21

Alignment Scores:
Pred. No.: 5,33e-68 Length: 888
Score: 668.00 Matches: 139
Percent Similarity: 63.0% Conservative: 49
Best Local Similarity: 46.64% Mismatches: 94
Query Match: 44.65% Indels: 16
DB: 9 Gaps: 4

US-10-062-624-40 (1-293) x US-10-059-964-21 (1-888)
Qy      6 LysPheThrIleAlaSerValLeuValCysLeuSerLeuProAsnIleSerSer 25
Db      10 AAATTTAAATTTGTAATGTTATTAACATTTTGTATTTCTTTTCCCACTTAAGTCA 69
Qy      26 -----SerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyrIleSerGly 43
Db      70 TTTACAACATATGCAAAATAAACAATCACTCAAAAAGTTGGATTGTACATAAGTGT 129
Qy      44 GlnTyrLysProSerValSerValPheSerAsnPheThrAlaValPheGlnAsp 63
Db      130 CAATAAAGCCCAAGTATCTCTCATTTCAAGAATTTTTCAGTAGAAGAAATGACAAAGTA 189
Qy      64 ThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLysThrAspAla 83
Db      190 GTA---GATTGTATGCTTTCACACTGATGTTACATATATACAGAACATATA----- 240
Qy      84 SerValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaValPheGlnAsp 103
Db      241 -----TTACGAGATAATACAAATTCACACCTCATTTATTTGCAAAAGTTCAAGAAC 291
Qy      104 AsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArgValGlu 123
Db      292 AAATTTATAAATTTACAGCAGTGCATTTGTTATTTATCTCGGCAAGGACCAAGGTTAGAA 351
Qy      124 IleGluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThrLeuSerAsp 143
Db      352 ATAGAAAGCTCTTATGGGATTTTGATGTTGTAATATATAAAATATATGTCAGTACAGAT 411
Qy      144 AlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLys--- 162
Db      412 GTTAATAGATAATTTTGTCTTAGTACGTGAAGAAATGTTTCAAAATTTCTCTCCAAACCA 471
```

Db 130 CAATATAGCCCAAGTATTCCTCATTTCAAGATTTTTCAGTAGAAGAAAATGACACAAAGTA 189  
Qy 64 ThrLysAsnLeuLeuAlaLeuLysLysAspValAspSerIleGluThrLysThrAspAla 83  
Db 190 GTA---GATTTCATAGGCTTACAACTGATGTTACATATACACAGAACAATATA----- 240  
Qy 84 SerValGlyIleSerAsnProSerAsnPherThrIleProTyrThrAlaValPheGlnAsp 103  
Db 241 -----TTACGAGATTAATACAAATTCACAACTCATTTATATTGCAAAAGTTTCAAGAAC 291  
Qy 104 AsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArgValGlu 123  
Db 292 AATTTTATAAATTCAGCAGTGCATTTGTTATTATCTGCGCAGGACCCAGGTTAGAA 351  
Qy 124 IleGluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThrLysSerAsp 143  
Db 352 ATAGAAAGCTCTTATGGGGAATTTGATGCTGTAATATAAAATTAATAAATATGTCAGTCAAGAT 411  
Qy 144 AlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLys--- 162  
Db 412 GTTAATAGATATTTTGTCTTATAGCTGAGGTAAGAAATGTTCCAAATTTCTCTCCAAACCA 471  
Qy 163 -----GluLysValSerAsnSerIlePheHisThrValMetArg 175  
Db 472 CATGAACCTAGTCAACCTCTGCACAGTAATCTCTAAAGAGTCTTTTATCTTTTATGAG 531  
Qy 176 AsnAspGlyLeuSerIleSerValIleValAsnValCysTyrAspPheSerLeuAsn 195  
Db 532 AATAATGGGTATTTGTCATCATGATAATATCAACGGTGTGTTATGATTTTCTTTTAAAT 591  
Qy 196 AsnLeuSerIleSerProTyrIleCysGlyAlaGlyValAspAlaIleGluPhePhe 215  
Db 592 AACACAAATATCACCTTACCTATGATATGATGAGGTTGGAGGAGATTTTATAGAGTTTTT 651  
Qy 216 AspValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuPro 235  
Db 652 GAAGTATGATATCAAGTTGCTGCGCAAGTAGGTTGGTATTTAGTATCTCCAAATCT 711  
Qy 236 SerAsnLeuSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLys 255  
Db 712 CCTCTATTACTATTTTGTCTGATGCAVATTTATCACAAGGTTCATAAATAATAATTTAAC 771  
Qy 256 AsnLeuAsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAlaVal 275  
Db 772 AACCTACATGTTAAGTATTCATATCACTTAAACCTCACCTACCATTCACCTCTGCAACA 831  
Qy 276 AlaThrLeuAsnIleGlyTyrPheGlyGlyGlyGlyAlaArgLeuThrPhe 293  
Db 832 CCCAACTAAACATTGAATATTTTGGTGGTGAAGTTGGGATGAGATTTATATTT 885

## RESULT 34

US-10-314-639-61  
; Sequence 61, Application US/10314639  
; Publication No. US20030103991A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/314,639  
; CURRENT FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US/09/314,701  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 61  
; LENGTH: 726  
; TYPE: DNA  
; ORGANISM: p30-12  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(726)

## US-10-314-639-61

Alignment Scores:  
Pred. No.: 1,26e-66 Length: 726  
Score: 655.00 Matches: 132  
Percent Similarity: 67.87% Conservative: 37  
Best Local Similarity: 53.01% Mismatches: 70  
Query Match: 43.78% Indels: 10  
DB: 9 Gaps: 4

## US-10-062-624-40 (1-293) x US-10-314-639-61 (1-726)

Qy 47 ProSerValSerValPheSerAsnPheSerValLysGluThrAsnValIleThrLysAsn 66  
Db 1 CCCGCTGTTTCTCATTTACAGTGACTTTTCAATTAAGAAACTTATCTAACTACAGGCA 60  
Qy 67 LeuIleAlaLeuLysLysAspValAspSerIleGluThrLysThrAspAlaSerValGly 86  
Db 61 TTGTTTGGGCTAAACCAAGATATTAGTTCTATTATTCGTAATAAAGAGACCCACA----- 114  
Qy 87 IleSerAsnProSerAsnPheThrIleProTyrThrAlaValPheGlnAsnSerVal 106  
Db 115 ---CAATATAATAACAATTTTAACGTTCCCTATCTACTGCAAAATTTCAAGACGACTTTCG 171  
Qy 107 AsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArgValGluIleGluGly 126  
Db 172 AGTTTCAGCATAGCTGTGATATATTTGCTAACCAATGCTCCAGAAATTTGAATAGAAGGA 231  
Qy 127 SerTyrGluGluPheAspValLysAsnProGlyGlyTyrThrLysSerAspAlaTyrArg 146  
Db 232 TCATTACGAAGAATTTGATGTTAAACCCAGGAATTTATACCAANTAGATGCTCATAGG 291  
Qy 147 TyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPhe---ThrProLysGluLysVal 165  
Db 292 TACATCTTTTAGCTAGAGAAAACCTTCTTACTATCTAAGTCTCTCTAAGAAAACAAA 351  
Qy 166 SerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleLysValIle 185  
Db 352 -----TATGTAATTTATAAAGAAATAACGGCATATCTATTGTATCTATTATA 396  
Qy 186 ValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSerProTyrIleCysGly 205  
Db 397 ATTAATGGTGTGATGATATTTCTTTAAATGATTCTAAGGTGTCCCTTACATATGCACA 456  
Qy 206 GlyAlaGlyValAspAlaIleGluPheAspValLeuHisIleLysPheAlaTyrGln 225  
Db 457 GGGTTTGGTGGAGATTTTAGAGTTTGTAGTCTATAGCTATAGCTTTTAAGTTTGTCTATCAA 516  
Qy 226 SerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSerLeuTyr 245  
Db 517 GGTAAAAATAGTATCATCTATTCATTATCTTCTAACATAATTTTATTACTGATGGATAT 576  
Qy 246 TyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGlnHisValAla---Glu 264  
Db 577 TACCACAAGGTAATAAATTCCTCAATTTTAAATTTTAAATGTTGAACATGTTGTTAAATGAG 636  
Qy 265 LeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIleGlyTyrPheGly 284  
Db 637 TTAACACTACAGATCCTCAAGTGACTTCTGCAACAGCATTTCTTAATATGATATTTGGT 696  
Qy 285 GlyGluIleGlyAlaArgLeuThrPhe 293  
Db 697 GGTGAATTTGGATTTAAATTTATATTT 723

## RESULT 35

US-10-059-964-61  
; Sequence 61, Application US/10059964  
; Patent No. US20020120115A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20020120115A1  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; TITLE OF INVENTION: Chaffeensis

```
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER:.. US/10/059,964
; CURRENT FILING DATE: 2002-01-28
; EARLIER APPLICATION NUMBER: 09/314,701
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61
; LENGTH: 726
; TYPE: DNA
; ORGANISM: p30-12
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(726)
US-10-059-964-61

Alignment Scores:
Pred. No.: 1,266-66 Length: 726
Score: 655.00 Matches: 132
Percent Similarity: 67.8% Conservatve: 37
Best Local Similarity: 53.01% Mismatches: 70
Query Match: 43.78% Indels: 10
DB: 12 Gaps: 4

US-10-062-624-40 (1-293) x US-10-059-964-61 (1-726)

Qy 47 ProSerValSerValPheSerAsnPheSerValLysGluThrAsnValIleThrLysAsn 66
Db ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 CCGCGTCTTCATACAGTGACCTTTCATTAAGAAACCTATATACACTAAGCACTGAGCA 60

Qy 67 LeuIleAlaLeuLysAspValAspSerIleGluThrLysThrAspAlaSerValGly 86
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 TTGTTGGCTAAACAGATATAGTCTATTTAGTAAAGAGACCA----- 114

Qy 87 IleSerAsnProSerAsnPheThrIleProTyrThrAlaValPheGlnAspAsnSerVal 106
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
115 ---CAATATAACAAATTTTAAACGTTCCCTATCTGCAAAATTTCAAGACGACITGCG 171

Qy 107 AsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArgValGluIleGluGly 126
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
172 AGTTTCAGCATAGCTGTTGGATATATTTGTAACAATGGTCCCAAGAAATTTGAAATAGAAGCA 231

Qy 127 SerTyrGluGluPheAspValLysAsnProGlyGlyTyrThrLysThrAspAlaTyrArg 146
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
232 TCTTACAGAAATTTGATGTTTAAACCCAGGAATTTATACAAATAGATGCTCATAGG 291

Qy 147 TyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPhe---ThrProLysGluLysVal 165
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
292 TACATTGCTTTAGTAGAGAAACCTTCTTACTATCTAAGTTCCTCTAAAGAAACAA 351

Qy 166 SerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSerValIle 185
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
352 -----TATGTAATTTATAAGAAATAACGGCATATCTATTCTATCTATTATA 396

Qy 186 ValAsnValCysTyrAspPheSerLeuAsnLeuSerIleSerProTyrIleCysGly 205
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
397 ATTATAGTTGTTATGATATTTCTTTAAATGATCTTAAGGTGTACCTTACATATGCACA 456

Qy 206 GlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIleLysPheAlaTyrGln 225
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
457 GGGTTGGTGGAGATTTTATAGATTTTTTATAGTCTATACGTTTAAAGTTGCTTATCAA 516

Qy 226 SerLysLeuGlyIleAlaTyrSerIleuProSerAsnIleSerIlePheAlaSerLeuTyr 245
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
517 GGTAAATAGGTATCAGTTATTTCATTATCTTCAACATAATTTTATTACTGATGATAT 576

Qy 246 TyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGlnHisValAla---Glu 264
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
577 TACCACAGGTAAATTTCCCAATTTAAATTTTAAATGTTGAACATGTTGTTAATGAG 636

Qy 265 LeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIleGlyTyrPheGly 284
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
637 TTAACCTACAGATCCTTAAAGTGACTTCTGCAACAGCAATTTCTTAAATATTTCAGTATTGCT 696
```

```
Qy 285 GlyGluIleGlyAlaArgLeuThrPhe 293
Db 697 GGTGAATTTGGATTAAATTTATATTT 723

RESULT 36
US-10-314-639-3
; Sequence 3, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; PRIOR FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/314,701
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 852
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(852)
US-10-314-639-3

Alignment Scores:
Pred. No.: 2,968-48 Length: 852
Score: 498.00 Matches: 112
Percent Similarity: 55.48% Conservatve: 60
Best Local Similarity: 36.13% Mismatches: 94
Query Match: 33.29% Indels: 44
DB: 9 Gaps: 11

US-10-062-624-40 (1-293) x US-10-314-639-3 (1-852)

Qy 1 MetAsnAsnLysLeuLysPheThrIleIleAsnThrValLeuValCysLeuSerLeu 20
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 ATGAATTACAAGAAATTTT---GTAAGCAGTGCAATTAATTTTCAATTAATGTCATC 54

Qy 21 -----ProAsnIleSerSerSerLysAlaIleAsnAsnAsn 32
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
55 TTACCTTACCAATCTTTTCAGATCCTGTAATCTTCAATGATACAGGAATCAACACAGC 114

Qy 33 AlaLysLysTyrTyrGlyLeuTyrIleSerGlyGlnTyrLysProSerValSerValPhe 52
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
115 AGAGAA-----GGCTTCTCATTAAGTAAAGTATAATCCNAGCATATCACACTTC 165

Qy 53 SerAsnPheSerValLysGluThrAsnVal-----IleThrLysAsnLeu 67
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
166 AGAAAAATTTTCAGCTGGAAGAGCTCCCATCAATGGAATACTTCTATCACTAAAAAGGT 225

Qy 68 IleAlaLeuLysLysAspValAspSerIleGluThrLysThrAspAlaSerValGlyIle 87
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
226 TTCGGCTGAAAAAGACGGAGAT-----ATA 252

Qy 88 SerAsnProSerAsnPheThrIleProTyrThrAlaVal---PheGlnAspAsnSerVal 106
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
253 GCACAACTCGGAATTTTAAACAGGACAGATCCAGCCCTCGAGTTTCAGAAATCACTATA 312

Qy 107 Asn---PheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArgValGluIleGlu 125
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
313 TCAGGATCTTCAGGAAGTATTTGGTTATGCTATG---GATGGCCCAAGAAATAGAACTTCAA 369

Qy 126 GlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThrLeuSer-----Asp 143
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
370 GCTGATACCAAAAAATTTGATGCAAAAAATCTTGACAAACATGACACTAAATACCGGTGAC 429

Qy 144 AlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGlu 163
```

```

Db 430 TACTATAAATCTTTGGACTATCTCGTGAA .....GACGCAATAGCAGATAAGAAA 480
Qy 164 LysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSer 183
Db 481 .....TATGTTGCTCTAAATGAAGGATGACCTTTTATGTC 519
Qy 184 VallileValAsnValCysTyrAspPheSerLeuAsnLeuSerIleSerProTyrIle 203
Db 520 TTAATGGTTAAACACTTGTATGACATATACGCTGAAGGAGTACCTTTTCATACCGTATGCA 579
Qy 204 CysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIleIlePheAla 223
Db 580 TGTGAGGTGTAGGAGCAGACCTTATAACGTAATTTAAGGATTTTAAATTTAAATTTCTCA 639
Qy 224 TyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSer 243
Db 640 TACCAAGGAAATAGGTATTAGCTATCAATCACACCAAGAGTTTCGCTTTTATTGGA 699
Qy 244 LeuTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGlnHisValAla 263
Db 700 GGATACTACCAGGAGTATTAGGAAATATTTTAAACAAATACCTGTATAACACCTGTA 759
Qy 264 GluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIleGlyTyrPhe 283
Db 760 GTATTAGAGGAGCTCTCAACAAACATCTGCGCTAGTAACTATTGACACTGGATACCTTT 819
Qy 284 GlyGlyGluIleGlyAlaArgLeuThrPhe 293
Db 820 GCGGAGAGTTGGAGTAAAGTTTCACCTTC 849

```

## RESULT 37

```

US-10-059-964-3
; Sequence 3, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikhisia, Yasuko
; APPLICANT: Ohasi, No. US20020120115A1
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; EARLIER APPLICATION DATE: 2002-01-28
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; LENGTH: 852
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(852)
US-10-059-964-3

```

## Alignment Scores:

```

Pred. No.: 2,96e-48 Length: 852
Score: 498.00 Matches: 112
Percent Similarity: 55.48% Conservative: 60
Best Local Similarity: 36.13% Mismatches: 94
Query Match: 33.29% Indels: 44
DB: 12 Gaps: 11

```

US-10-062-624-40 (1-293) x US-10-059-964-3 (1-852)

```

Qy 1 MetAsnLysLeuLysPheThrIleLeuAsnThrValLeuValCysLeuLeuSerLeu 20
Db 1 ATGATTACAAGAAATTTT .....GTAAGCAGTGCATTAATTTCAATTAATGTCATC 54
Qy 21 .....ProAsnIleSerSerLysAlaIleAsnAsnAsn 32
Db 55 TTACCTTACCAATCTTTTGGAGATCCCTGTAACCTTCAATATGATACAGGAATCAACGACAGC 114

```

```

Qy 33 AlaLysLysTyrTyrGlyLeuTyrIleSerGlyGlnTyrLysProSerValSerValPhe 52
Db 115 AGAGAA .....GGCTTCTACATTAGTGTAAAGTATATCCAGCATATCACACTTC 165
Qy 53 SerAsnPheSerValLysGluThrAsnVal .....IleThrLysAsnLeu 67
Db 166 AGAAATTTCTCAGCTGAAGAGCTCCCATCAATGAAATACTTTCTATCATAAAAGGTT 225
Qy 68 IleAlaLeuLysAspValAspSerIleGluThrLysThrAspAlaSerValGlyIle 87
Db 226 TTCGGCTGAAAAAAGCGGAGAT .....ATA 252
Qy 88 SerAsnProSerAsnPheThrIleProTyrThrAlaVal .....PheGlnAspAsnSerVal 106
Db 253 GCACAAATCTGCGAATTTTAAACAGGACAGATCCAGCCCTCGAGATTTTCAGATAACCTAATA 312
Qy 107 Asn .....PheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArgValGluIleGlu 125
Db 313 TCAGGATTTCTCAGGAAGTATTGGTTATGCTATG .....GATGGGCCAAGAAATAGAACTTGA 369
Qy 126 GlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThrLeuSer .....Asp 143
Db 370 GCTGCATACCAAAATTTGATGCAAAATCTGCACAAATGACACTAATAGCGGTGAC 429
Qy 144 AlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGlu 163
Db 430 TACTATAAATCTTTGGACTATCTCGTGAA .....GACGCAATAGCAGATAAGAAA 480
Qy 164 LysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSer 183
Db 481 .....TATGTTGCTCTTAAAAATGAAGCATCCTACTTTTATGTCA 519
Qy 184 VallileValAsnValCysTyrAspPheSerLeuAsnLeuSerIleSerProTyrIle 203
Db 520 TTAATGGTTAAACACTTGTCTATGACATTACACTGAAGGAGTACCTTTTCATACCGTATGCA 579
Qy 204 CysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIleLysPheAla 223
Db 580 TGTGAGGTGTAGGAGCAGACCTTATAACGTAATTTAAGGATTTTAAATTTAAATTTCTCA 639
Qy 224 TyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSer 243
Db 640 TACCAAGGAAATAGGTATTAGCTATCAATCACACCAAGAGTTTCGCTTTTATTGGA 699
Qy 244 LeuTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGlnHisValAla 263
Db 700 GGATACTACCAGGAGTATTAGGAAATATTTTAAACAAATACCTGTATAACACCTGTA 759
Qy 264 GluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIleGlyTyrPhe 283
Db 760 GTATTAGAGGAGCTCTCAACAAACATCTGCGCTAGTAACTATTGACACTGGATACCTTT 819
Qy 284 GlyGlyGluIleGlyAlaArgLeuThrPhe 293
Db 820 GCGGAGAGTTGGAGTAAAGTTTCACCTTC 849

```

## RESULT 38

```

US-10-062-051-41
; Sequence 41, Application US/10062051
; Publication No. US20030073095A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/10/062,051
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US/09/660,587
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358

```





Db 328 TACTCTATG---GACGGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCA 384  
Qy 134 LysAsnPro-----GlyGlyTyrThrLeuSerAspAlaTyrArg 146  
Db 385 AAAAACCCGATAAACAATGATACTGATAATGCTGAATAC-----TATAAA 429  
Qy 147 TyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGluLysValSer 166  
Db 430 CATTTTGATATATCTCGT-----AAAGATGCAATGGAA 462  
Qy 167 AsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSerValIleVal 186  
Db 463 GATCAGCAATAT---GTAGTACTTAAATGACGCAATCTTTATGTCATGATGTT 519  
Qy 187 AsnValCysTyrAspPheSerLeuAsnLeuSerIleSerProTyrIleCysGlyGly 206  
Db 520 AATACCTGCTATGATACATTAAGTGAAGAGTATCTTTTCGTACCATATGTCATGTCAGGT 579  
Qy 207 AlaGlyValAspAlaIleGluPheAspValLeuHisIleLysPheAlaTyrGlnSer 226  
Db 580 ATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGA 639  
Qy 227 LysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSerLeuTyr 246  
Db 640 AAAATAGGTATTAGTATCTTACCTATCACCAGAGTCTCTGCATTTATTTGGTGATACTAC 699  
Qy 247 HisLysValMetGlyAsnGlnPheLysAsnValGlnHisValAlaGluLeuAla 266  
Db 700 CATGCGTTATTGGTAAATAATTTGAGAAGATACCTGTAATACTCTGTTAGTATTAAAT 759  
Qy 267 SerIleProLysIleThrSerAlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGlu 286  
Db 760 GATGCTCTCAAAACACATCTGCTTCAGTAACTCTTGAGTGGATACTTTGGCGGAGAA 819  
Qy 287 IleGlyAlaArgLeuThrPhe 293  
Db 820 ATTGGAAATGAGGTTTCACTTC 840

## RESULT 40

US-10-062-624-41  
; Sequence 41, Application US/10062624  
; Patent No. US20020115840A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-Kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2/D1  
; CURRENT APPLICATION NUMBER: US/10/062,624  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: 09/660,587  
; PRIOR FILING DATE: 2000-09-12  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 41  
; LENGTH: 840  
; TYPE: DNA  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: nucleic acid sequence of E. canis p28-2  
US-10-062-624-41

Alignment Scores:  
Pred. No.: 2,16e-47 Length: 840  
Score: 490.50 Matches: 114  
Percent Similarity: 56.35% Conservative: 59  
Best Local Similarity: 37.13% Mismatches: 85  
Query Match: 32.79% Indels: 49  
DB: 12 Gaps: 12

US-10-062-624-40 (1-293) x US-10-062-624-41 (1-840)

Qy 6 LysPheThrIleIleAsnThrValLeuValCysLeuLeuSer---LeuProAsnIleSer 24

Db 10 AAGAAAAATTCCTAGTAAGAAGCGCGTTAATCTTCATTAATGTCAATCTTACCATATCATGCT 69  
Qy 25 SerSerLysAlaIle-----AsnAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40  
Db 70 TTTCAGATCCTGTAGGTTCAAGAACTAATGATAAACAAGAA-----GGCTTCTAC 120  
Qy 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGlu--- 59  
Db 121 ATTAGTCAAAAGTACAATCCAAGTATATCACACTTTAGAAAATTTCTCTGCTGAAGAAACT 180  
Qy 60 -----ThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAsp 75  
Db 181 CCTATTAAATGAACAAATTTCTCTCACTAAAAAAGTTTTCGACACTAAAGAAAGATGGTGAT 240  
Qy 76 SerIleGluThrLysThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIle 95  
Db 241 -----ATACAAAAAAGAGCATTTTACAGA 267  
Qy 96 ProTyrThrAlaVal---PheGlnAspAsnSerValAsn---PheAsnGlyThrIleGly 113  
Db 268 GTAGCTCCAGCAATGATTTTCAAAATAACTTAATATCAGGATTTTCAGGAAGTATTGGT 327  
Qy 114 TyrThrPheAlaGluGlyThrArgValGluIleGluGlySerTyrGluGluPheAspVal 133  
Db 328 TACTCTATG---GACGGACCAAGAAATAGAACTTGAAGCTGCATATCAACAATTTAATCCA 384  
Qy 134 LysAsnPro-----GlyGlyTyrThrLeuSerAspAlaTyrArg 146  
Db 385 AAAAACCCGATAAACAATGATACTGATAATGCTGAATAC-----TATAAA 429  
Qy 147 TyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGluLysValSer 166  
Db 430 CATTTTGATATATCTCGT-----AAAGATGCAATGGAA 462  
Qy 167 AsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSerValIleVal 186  
Db 463 GATCAGCAATAT---GTAGTACTTAAATGACGCAATCTTTATGTCATGATGTT 519  
Qy 187 AsnValCysTyrAspPheSerLeuAsnLeuSerIleSerProTyrIleCysGlyGly 206  
Db 520 AATACCTGCTATGATACATTAAGTGAAGAGTATCTTTTCGTACCATATGTCATGTCAGGT 579  
Qy 207 AlaGlyValAspAlaIleGluPheAspValLeuHisIleLysPheAlaTyrGlnSer 226  
Db 580 ATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGA 639  
Qy 227 LysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSerLeuTyr 246  
Db 640 AAAATAGGTATTAGTATCTTACCTATCACCAGAGTCTCTGCATTTATTTGGTGATACTAC 699  
Qy 247 HisLysValMetGlyAsnGlnPheLysAsnValGlnHisValAlaGluLeuAla 266  
Db 700 CATGCGTTATTGGTAAATAATTTGAGAAGATACCTGTAATACTCTGTTAGTATTAAAT 759  
Qy 267 SerIleProLysIleThrSerAlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGlu 286  
Db 760 GATGCTCTCAAAACACATCTGCTTCAGTAACTCTTGAGTGGATACTTTGGCGGAGAA 819  
Qy 287 IleGlyAlaArgLeuThrPhe 293  
Db 820 ATTGGAAATGAGGTTTCACTTC 840

Search completed: July 4, 2003, 00:12:28  
Job time : 154 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 3, 2003, 22:55:53 ; Search time 1091 Seconds  
(without alignments)  
4349.475 Million cell updates/sec

Title: US-10-062-624-40  
Perfect score: 1496  
Sequence: 1 MNKMKFTIINTVLVCLLSL.....AVATLNIGYGGEGIGARLTF 293

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10062624/runat\_30062003\_091229\_24315/app\_query.fasta\_1.455  
-DB=EST -QFMT=FASTAP -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10062624.cgn 1 1456 @runat\_30062003\_091229\_24315 -NCPU=6 -ICPU=3  
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	113.5	7.6	842	17	AZ670804 ENTMU46TF
2	101	6.8	719	10	AW773357 IP03f06.Y
C 3	100.5	6.7	997	17	AZ678995 ENTIG68TR
4	98.5	6.6	700	9	AL507611 AL507611
5	97.5	6.5	871	9	AL667424 AL667424
6	97	6.5	792	13	BI603807
7	96.5	6.5	649	13	BJ329029 BJ329029
8	96.5	6.5	728	14	BP001170 BP001170
9	95.5	6.4	837	9	AL669478 AL669478
10	95.5	6.4	862	9	AL669078 AL669078
11	94.5	6.3	648	10	AV938117 AV938117
12	93.5	6.2	547	9	AI224977 AI224977
C 13	92.5	6.2	753	17	AA550337 AA550337
14	92.5	6.2	880	17	AZ534829 ENTCK35TF
15	92	6.1	584	13	BJ426331 BJ426331
16	92	6.1	594	13	BJ420538 BJ420538
17	92	6.1	595	13	BJ336589 BJ336589
18	92	6.1	599	13	BJ427735 BJ427735
19	92	6.1	625	13	BJ426166 BJ426166
20	92	6.1	641	13	BJ424116 BJ424116
21	92	6.1	644	13	BJ420225 BJ420225
22	92	6.1	645	13	BJ418961 BJ418961
23	92	6.1	702	12	EG594518 EG594518
24	91.5	6.1	627	13	BJ415153 BJ415153
25	91.5	6.1	649	13	BJ410964 BJ410964
26	91.5	6.1	650	13	BJ412332 BJ412332
27	91.5	6.1	918	9	AL668855 AL668855
28	91	6.1	584	13	BJ423418 BJ423418
29	91	6.1	590	13	BJ332457 BJ332457
30	91	6.1	590	13	BJ424236 BJ424236
31	91	6.1	591	13	BJ422368 BJ422368
32	91	6.1	591	13	BJ425564 BJ425564
33	91	6.1	591	13	BJ425783 BJ425783
34	91	6.1	592	13	BJ425553 BJ425553
35	91	6.1	594	13	BJ329215 BJ329215
36	91	6.1	596	13	BJ423262 BJ423262
37	91	6.1	597	13	BJ333160 BJ333160
38	91	6.1	597	13	BJ335009 BJ335009
39	91	6.1	597	13	BJ418039 BJ418039
40	91	6.1	597	13	BJ422231 BJ422231
41	91	6.1	598	13	BJ418067 BJ418067
42	91	6.1	599	13	BJ419059 BJ419059
43	91	6.1	607	13	BJ427530 BJ427530
44	91	6.1	609	13	BJ332974 BJ332974
45	91	6.1	611	13	BJ419336 BJ419336

# ALIGNMENTS

RESULT 1  
AZ670804/c  
LOCUS  
DEFINITION AZ670804 Entamoeba histolytica 842 bp DNA linear GSS 14-DEC-2000  
GENOMIC, DNA sequence.  
ACCESSION AZ670804.1 GI:11807950  
VERSION GSS.  
KEYWORDS Entamoeba histolytica.  
SOURCE Entamoeba histolytica.  
ORGANISM Eukaryota; Entamoebidae; Entamoeba.  
REFERENCE 1 (bases 1 to 842)  
AUTHORS Loftus, B., Van Aken, S. and Fraser, C.  
TITLE Determination of clone end sequences from Entamoeba histolytica  
HM1:IMSS sheared DNA library

JOURNAL  
COMMENT

Unpublished (2000)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjoftus@igr.org  
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared  
DNA library  
Seq primer: M13-Forward  
Class: shotgun  
High quality sequence start: 18  
High quality sequence stop: 798.  
Location/Qualifiers  
1. .842  
/organism="Entamoeba histolytica"  
/strain="HM1:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"  
/note="Vector: pGSI; Site1: Bst I; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450.). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + i method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barell, Oxford University Press, 1999)."

FEATURES  
source

BASE COUNT 303 a 106 c 97 g 336 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 0.00147 Length: 842  
Score: 113.50 Matches: 67  
Percent Similarity: 37.54% Conservative: 40  
Best Local Similarity: 23.51% Mismatches: 75  
Query Match: 7.59% Indels: 103  
DB: 17 Gaps: 15

US-10-062-624-40 (1-293) x AZ670804 (1-842)

QY 25 SerSerLysAlaIleAsnAsnAlaLysLysTyrGlyLeuTyrIleSer----- 42  
Db 837 TCCATAAAATTTTATTATCCCTTGTCTAAATAA---TATGGAGTTATATTAGTCAGCAC 781  
QY 43 GlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThrAsnVal 62  
Db 780 GCACAGTATTTCATGGCGGTTTCTTCAATTTCAAAATGTAGATTATCA----- 733  
QY 63 IleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLysThrAsp 82  
Db 732 -----AACTTCCTTTGTCTATTATTGATTAAACACCA 697  
QY 83 AlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaVal---Phe 101  
Db 696 TTAATAATTAATAACGAATAATGAACATATATATCTCTTTATTATCAATTTGTTAAAGTA 637  
QY 102 GlnAspAsnSerValAsn-----PheAsnGlyThrIleGlyTyrThr 115  
Db 636 GAACCTAATTTTAAATGAAGTATCAACAACTCTTTTAAAGGAATGAGTGTATTGTA 577  
QY 116 PheAlaGlu---GlyThrArgValGluIleGluGlySerTyrGluGluPheAspValLys 134  
Db 576 ATGAATGAAGAAACAACTAACCTTTTATCTCTTTTGAACAAACGATGTTAT 517  
QY 135 AsnProGlyGlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMet 154  
Db 516 AAT-----CTTAATGATTGTTTAAAAAATTT----- 490

QY 155 LysGlyAsnSerPhe-----ThrProLys-----Glu 163  
Db 489 GATTCTAACTCTTATCTATTGTTGTTCTTCAACAAACCAAAATTTTGGCAACAC 430  
QY 164 LysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSer 183  
Db 429 GAAGTAGCAGCAAGTTTACACCAATCAATAGTATTACATGATGACACATAGATTAAATGGA 370  
QY 184 VallleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSerProTyrIle 203  
Db 370 ----- 370  
QY 204 CysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIleLysPheAla 223  
Db 369 -----TCATTTAAATATTGTTCAATG----- 346  
QY 224 TyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSer 243  
Db 345 -----AAT 343  
QY 244 LeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsn-----LeuAsn 258  
Db 342 TTAATTTTATATAAAGAAATTTTATTAAATCAATTTTAAACCTTTGAATGTAAACAAATTAAT 283  
QY 259 ValGlnHisVal---AlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThr 277  
Db 282 ATAGAAATATTAATTAATCAATTTAGATTCTTTTAAATATAACTCATCAA---AGAACA 226  
QY 278 LeuAsnIleGlyTyr 282  
Db 225 GAAATATTCTTTAT 211

## RESULT 2

AW773357

## LOCUS

AW773357

## DEFINITION

xp03f06.y1 Kloeak Brook Zeldia punctata SL1 Zeldia punctata cDNA 5'

similar to WP:TI0B5.7 CE18237 LIPASE ;, mRNA sequence.

## ACCESSION

AW773357

## VERSION

AW773357.1 GI:7710318

## KEYWORDS

EST.

## SOURCE

Zeldia punctata.

## ORGANISM

Zeldia punctata

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;

Cephaloboidae; Cephalobidae; Zeldia.

1 (bases 1 to 719)

McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,

Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,

Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagaris, V., R.,

Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe

, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.

, Shin, T., Jackson, F., Cardenas, M., McCann, R., Waterston, R. and

Wilson, R.

The Washington Univ. Nematode EST Project, 1999

Unpublished (1999)

Contact: McCarter JP

The Washington Univ. Nematode EST Project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

The library was constructed by Merry Brook and Dr. Andrew Kloeak

(kloeak@biology.wustl.edu) at Divergence LLC. DNA Sequencing by:

Washington University Genome Sequencing Center St. Louis.

Seq primer: SL1 primer

High quality sequence stop: 518.

## FEATURES

source

1. .719

/organism="Zeldia punctata"

/strain="PDL3"

/db\_xref="taxon:49351"

/clone\_lib="Kloeak Brook Zeldia punctata SL1"



```

Qy 21 ProAsnIleSerSerSerLysAlaIleAsnAsnAlaLysLysTyrGlyLeuTyr 40
Db 640 GGTAAATTTAAATCT- - - - - GTTACAATAACGGTAAAGAGACACAGGTTATGGT 590
Qy 41 IleSerGlyGlnTyrLysProSerValPheSerAsnPheSerValLysGluThr 60
Db 589 ATTGGAATTAATATAATCCAGCAAACTCACTATATATATGCTTAATTCAGCATATAT 530
Qy 61 AsnValIleThrLysAsnLeuLeuAlaLeuLysLysAspVal- - - - - AspSerIleGlu 78
Db 529 GGAGTAACCTGGTCTTAATGTTGGAATTTGGAGGTAATGTTAAACACGATAAAGTACAT 470
Qy 79 ThrLysThrAspAlaSerValGlyLysLeuAsn- - - - - ProSerAsnPheThrIleProTyr 97
Db 469 CTTAGTCTAATGCTGAGTTGGAAAGAAATACAAAAGCGCAGCTTACCTTTCATAT 410
Qy 98 ThrAlaValPheGlnAspAsnValAsnPheAsnGlyThrIleGlyTyrThrPheAla 117
Db 409 AAT- - - - - CCATCTGATATAATCTCTGTTGATGTTAAAGTTGGTTATCTTATGCT 359
Qy 118 GluGly- - - - - ThrArgValGluIleGluGlySerTyr- - - - - GluGluPhe 131
Db 358 AATGACATGATCTTCTTAATTAACGAGTGTGATTTACACATAACACAGATCAAAAGTGG 299
Qy 132 AspValLysAsnProGlyTyrThrLysSerAspAlaTyrArgTyrPheAlaLeuAla 151
Db 298 GATATTTCT- - - - - - - - - - - GCTGGTGTCTTCT 278
Qy 152 ArgGluMetLysGlyAsnSerPheThrProLysGluLysValSerAsnSerIlePheHis 171
Db 277 CGTTCAATGAAGGAAGTACTTGGAGTGCTAAAGCAGCA- - - - - CAT 236
Qy 172 ThrValMetArgAsnAspGlyLeuSerIleLeuSerValIleValAsnValCysTyrAsp 191
Db 235 AAA- - - - - ATGAATAAAGACTCTTCGTGGTTTCATTAGT- - - - - 203
Qy 192 PheSerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAsp- - - 210
Db 202 - - - - - GCTGAAAAGGACCATCAGGATCCAGATATATGGAAGGTTTCAAAATATCCTTT 146
Qy 211 - - - - - AlaIleGluPhePheAspValLeuHisIleLysPheAlaTyrGln 225
Db 145 TAATTAATTCGACTTCTTCTTCAATT- - - - - CTTTATTAATTCGATTGTTATCAG 92
Qy 226 SerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPhe 241
Db 91 AAATAACTTTAAATAAAA- - - - - ACACCTTTCCAGCCACAAAAGCCTTTTC 47

RESULT 4
AL507611 700 bp mRNA linear EST 04-JAN-2001
DEFINITION AL507611 Hordeum vulgare Barke developing caryopsis (3.-15.DAP)
            Hordeum vulgare cDNA clone HY06F12V 5', mRNA sequence.
ACCESSION AL507611
VERSION AL507611.1 GI:12033826
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
          ; Triticeae; Hordeum.
          1 (bases 1 to 700)
REFERENCE Michalek, W., Weschke, W., Pleisner, K.-P. and Graner, A.
AUTHORS EST sequencing and analysis in barley
TITLE Unpublished (2000)
JOURNAL Contact: Michalek W.
COMMENT Institute for Plant Genetics and Crop Plant Research
          Corrensstr. 3, D-06466 Gatersleben, Germany
          Email: michalek@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de
          Seq primer: r3 primer for 5' end.
          Location/Qualifiers
            1..700
              /organism="Hordeum vulgare"

```

```

/cultivar="Barke"
/db_xref="taxon:4513"
/clone="HY06F12V"
/clone_lib="Hordeum vulgare Barke developing caryopsis
(3.-15.DAP)"
/tissue_type="developing caryopsis (3.-15.DAP)"
/lab_host="XL0LR"
/notes="vector: plasmid pBK-CMV; Site_1: EcoRI; Site_2:
XhoI; mRNA was made from developing caryopsis (3.-15.DAP)
of spring barley variety 'Barke', a high quality malting
variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI
(3'-end of cDNA). NOTE: Due to a cloning artefact caused
by the kit, in most cases the EcoRI site is NOT present,
as well as the EcoRI adapter. Average insert size is 1 kb
Sequence trimming: Vector sequences and sequence ends were
trimmed from the 5'- and 3'-end until a 50 bp window
contains less than two ambiguities. The maximum length was
set to 700 bp"
BASE COUNT 185 a 154 c 176 g 183 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 0.0907 Length: 700
Score: 98.50 Matches: 50
Percent Similarity: 36.04% Conservative: 21
Best Local Similarity: 25.38% Mismatches: 101
Query Match: 6.58% Indels: 25
DB: Gaps: 8
US-10-062-624-40 (1-293) x AL507611 (1-700)
Qy 40 TyrIleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGlu 59
Db 117 TACATCATCGCGCATTCACAGCCGCCCTCGACATCTTCGTGACCTTTCCGATGAGAGG 176
Qy 60 ThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAsp- - - - - 73
Db 177 - - - - - TCCCGAGACAGGTTGCAGTCAAGAGGATATATGGAGACGCCGATG 224
Qy 74 - - - - - ValAspSerIleGluThrLysThrAspAlaSerValGlyIleSerAsnPro 90
Db 225 GTGCGCGGTTCCAGAGCGCTGGAGACC- - - ATAGCCGAGAGAGGTATCAATAGCGCTGT 281
Qy 91 SerAsnPheThrIleProTyrThrAlaValPheGlnAsnAsnSerValAsnAsnGly 110
Db 282 CCTGGTAAAGGATTGACACATGCGGTGT- - - - - AAGATTGAGCTGCTGCT 329
Qy 111 ThrIleGlyTyrThrPheAlaGluGlyThrArgValGluIleGluGlySerTyrGluGlu 130
Db 330 CAGATAGAGTTGATTTTCGACAGAGGCAACTTCTACGACTTNACTTCAATTTGGTGGTGAG 389
Qy 131 PheAspValLysAsnProGlyGlyTyrThrLysSerAspAlaTyrArgTyrPheAlaLeu 150
Db 390 TTAGATGTT- - - - - CCTGGTGAATATATCAAGAAAGACATATCCATTTGAGTTCTCT 443
Qy 151 AlaArgGluMetLysGlyAsnSerPheThrProLysGluLysValSerAsnSerIlePhe 170
Db 444 ACTGTTGAAATGCCATATGATGATCATACATGGAACAAATGTACAGCTGAGTATCTCTG 503
Qy 171 HisThrValMetArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysTyr 190
Db 504 AAAGTGACGATT- - - - - GGTAGAACTATGTTGGCAATATTTGGAA- - - TGTCGG 551
Qy 191 AspPheSerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAsp 210
Db 552 GATTCTGTGTAGGAACACTATCTCCGCTTCCTTCGATCAATAACAGCATCAAGATGAA 611
Qy 211 AlaIleGluPheAspValLeuHisIleLysPheAlaTyrGlnSerLys 227
Db 612 - - - GTTGGATTGAAGATTGTCTGCAATATTGATTGATACAGCAAAAAG 659

RESULT 5
AL667424

```

```

LOCUS      AL667424                871 bp    mRNA    linear    EST 11-JAN-2002
DEFINITION Clone 0192D04 5', mRNA sequence.
ACCESSION  AL667424
VERSION    AL667424.1  GI:18134331
KEYWORDS  EST.
SOURCE    Ciona intestinalis.
ORGANISM  Ciona intestinalis
           Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
           Phlebobranchia; Cionidae; Ciona.
REFERENCE  1 (bases 1 to 871)
           Ciona intestinalis directional larval cDNA library
TITLE      Ciona intestinalis directional larval cDNA library
JOURNAL    Unpublished (2002)
COMMENT    Contact: Genoscope
           Genoscope - Centre National de Sequencage
           BP 191 91006 EVRY cedex - France
           Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
           IMPORTANT: this sequence may contain errors. The Ciona intestinalis
           library from which the clone was isolated may be contaminated with
           cDNAs from bacteria or other Eukarya.
           Directional larval cDNA library originate from Dr M.Branno,
           Stazione A.Dohrn, Naples, Italy, and was prepared in
           pBluescript2SK+.
FEATURES   Location/Qualifiers
             1..871
             /organism="Ciona intestinalis"
             /db_xref="taxon:7719"
             /clone="0192D04"
             /clone_lib="directional larval cDNA library"
             /note="Vector: pBluescript2SK+"
BASE COUNT 249 a 199 c 196 g 226 t
ORIGIN
Alignment Scores:
Pred No.:      0,179      Length:      871
Score:         97.50      Matches:    55
Percent Similarity: 38.49%      Conservative: 47
Best Local Similarity: 20.75%      Mismatches: 102
Query Match:    6.52%      Indels:    61
DB:             9         Gaps:      10

US-10-062-624-40 (1-293) x AL667424 (1-871)

Qy      5 LeuLysPheThrIleAlaAsn---ThrValLeuValCysLeuLeuSerLeuProAsnIle 23
Db      31 CTAAAGTTCAACATGATGAAGTGACGGTTATGCACTCTGCTGTTATCGACCAAGCC 90
Qy      24 SerSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyrIleSerGly 43
Db      91 ATTGAGCGACATGGGTGTACAGCCAGCAAGCAAAAGGGCCTTCCAAATTGGCTGAATTC 150
Qy      44 GlnTyrLysProSerValSerValPheSerAsnPheSerValLys---GluThrAsnVal 62
Db      151 TAGCCCGCTTCGGGCACAAAGCGCAATCCAAATCAGCATCGACATGAAATTCGGTA 210
Qy      63 IleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLysThrAsp 82
Db      211 GCCAATAAGAACCTTGGAAACGTTTGATCGG-----TCGGTGTTCAGTCACTTCCT 261
Qy      83 AlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaValPheGln 102
Db      262 GCCTCGATGGAGTGGTTAAAT-----
Qy      103 AspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArgVal 122
Db      283 -----AATGGA-----CATGCATTGCAAGTC 303
Qy      123 GluIleGluGlySerTyrGluGluPheAspValLys---AsnProGlyGlyTyrThrLeu 141
Db      304 AACATGGACGATCTTACACATATCTGATCGAAGTGGCTCCCAATGATATNAGGCT 363
Qy      142 SerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrPro 161

```

```

Db      364 GTTCAGTTCCATCTTCACTGGCTGCTGCCAACAGAGCGAAGGTTCTGAGCATTTGGCTT 423
Qy      162 LysGluLysValSerAsnSerIlePheHisThrValMetArgAsn----- 176
Db      424 AACGGCAAGCATACTTCCGAGAGTACATGTTGTTTCATTATATATACTAAGTATGCTAGC 483
Qy      177 -----AspGlyLeuSerIleLeuSerValIleValAsnVal 188
Db      484 ATTGGGGAAGCGGTAAATAAGCCAGATGGCTCCAGCTACTTGTGTCTTTGTTGATATC 543
Qy      189 CysTyrAspPheSerLeuAsn---AsnLeuSerIleSerProTyrIleCysGlyAla 207
Db      544 AATGACGAATCAACAGGACATTTTAATCTTTTGTCTAAACCGG----- 585
Qy      208 GlyValAspAlaIleGluPhePheAspValLeuHisIleLysPheAlaTyrGlnSerLys 227
Db      586 -----ATCAACCGTGTCAATAATCAACCAAGCAAGAAATGACGTATACATCAAG 633
Qy      228 LeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrHis 247
Db      634 TTTTCGTCGCAAAATCTTTGCCAACAGACTTGTC-----GAATATTAC 678
Qy      248 LysValMetGlyAsn 252
Db      679 AGATACAGGGGATCA 693

RESULT 6
LOCUS    BI603807
DEFINITION 603244678F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5286979 5',
           mRNA sequence.
ACCESSION BI603807
VERSION   BI603807.1  GI:15496747
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 792)
           NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE     Unpublished (1999)
JOURNAL   Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-remail.nih.gov
           Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
           cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
           Toshituki and Piero Carninci (RIKEN)
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LUAM11725 row: h column: 20
           High quality sequence stop: 769.
FEATURES  Location/Qualifiers
           1..792
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
           /clone="IMAGE:5286979"
           /clone_lib="NIH_MGC_96"
           /tissue_type="hypothalamus"
           /lab_host="DH10B"
           /note="Organ: brain; Vector: pBluescriptR (modified
           pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
           ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
           size-selected for average insert size 2.3 kb and
           normalized to ROT 5. This is a primary library enriched
           for full-length clones and constructed using the
           Cap-trapper method (Carninci, in preparation). Library
           constructed by M. Brownstein (NIH/NHGRI, National
           Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 275 a 129 c 154 g 234 t

```

```

ORIGIN
Alignment Scores:
Pred. No.: 0.176 Length: 792
Score: 97.00 Matches: 59
Percent Similarity: 38.40% Conservative: 32
Best Local Similarity: 24.89% Mismatches: 98
Query Match: 6.48% Indels: 48
DB: 13 Gaps: 11

US-10-062-624-40 (1-293) x BJ603807 (1-792)
QY 2 AsnAsnLysLeuLysPheThrIleLeuThr-----ValLeuValCysLeuLeu 18
DB 55 AACATCAAACTAAGAAATGCAAAATGAATCTCTTCTATATATATTAATACTAATCTTTT 114
QY 19 SerLeuProAsn-----IleSerSerSerLysAlaIleAsnAsn---AsnAla 33
DB 115 TCTTTGTTTCTCAAGGTATTTTACTTTCAGCATCCAGTCCATGAAGAAATTTAGATGAT 174
QY 34 LysLysTyTyTyGlyLeuTyTyIleSerGlyGlnTyTyLysProSerValSerValPheSer 53
DB 175 GACATGGTATTATATACATTACAGTTGGGAAA-----GGCTTTTCAG 216
QY 54 AsnPheSerValLysGluThrAsnValIleThrLysAsnLeuLeuAlaLeuLysLysAsp 73
DB 217 AAGGAGACACTGCAGAAATATGCTTCTCTCGGACAAATATATAAATGAT 276
QY 74 ValAspSerIle-----GluThrLysThrAspAlaSerValGlyIleSerAsn 89
DB 277 GAGACAGTTTCATGAACGAGGAAATAAAGTTTCAAGAACACACAGGCTCCAAACAT 336
QY 90 -----ProSerAsnPheThrIle---ProTyThrAlaValPhe 101
DB 337 AATTTCCTTAAATCATGGTCTGCCACTGAATCTGGCTATAAAACCTTATCTTGCACTA--- 393
QY 102 GlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyThrPheAlaGluGlyThrArg 121
DB 394 ---AAGGATCTGTAGCTTCCAGCTGAGAAATGGAGTTCAGAACTGAATCAACACAA 450
QY 122 ValGluIleGluGlySerTyTyGluGluPheAspValLysAsnProGlyGlyTyThrLeu 141
DB 451 GAAAGAGAGAAATTTGGGGATGAAGAAACTCAGCTAAATTTCTTAGTA----- 501
QY 142 SerAspAlaTyArgTyTyPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrPro 161
DB 502 -----AGGAGATTTTGACATGCTCAGATGTATGCTGGGAGAGCTCTACCGACCT 552
QY 162 LysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSer--- 180
DB 553 TGTGGCAAGTCTGATACCTGTTGGTCCACATCATCTTTTTCAGAGAAATAAAGCATT 612
QY 181 -----IleIleSerValIleValAsnVal 188
DB 613 TAATTGCCAATGGAGAGAGCCCATCTACTACTATACTTGTGTTATGTTAATGTC 672
QY 189 CysTyAspPheSer---LeuAsnAsnLeuSerIleSerProTyTyIleCys 204
DB 673 TGTTTTAAAGAAAGTAGTGTGTTAAGATGTATCAGTAATCAGTAATGATATGTC 723

RESULT 7
BJ329029 649 bp mRNA linear EST 05-MAR-2002
LOCUS BJ329029 Dictyostelium discoideum cDNA library, AF Dictyostelium
DEFINITION dictyostelium cDNA clone dda30c13 5', mRNA sequence.
ACCESSION BJ329029
VERSION BJ329029.1 GI:19159159
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
AUTHORS Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.

```

Full length cDNA of Dictyostelium discoideum at the aggregation stage

Unpublished (2002)  
Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.

Location/Qualifiers  
1..649  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone="dda30c13"  
/sex="mat A"  
/dev\_stage="aggregation stage"  
272 a 112 c 82 g 183 t

BASE COUNT 272 a 112 c 82 g 183 t

ORIGIN

Alignment Scores:  
Pred. No.: 0.144 Length: 649  
Score: 96.50 Matches: 60  
Percent Similarity: 35.07% Conservative: 34  
Best Local Similarity: 22.39% Mismatches: 90  
Query Match: 6.45% Indels: 84  
DB: 13 Gaps: 12

US-10-062-624-40 (1-293) x BJ329029 (1-649)

QY 21 ProAsnIleSerSerSerLysAlaIleAsnAsnAsnAlaLysLysTyTyTyTyGlyLeuTy 40  
DB 31 CCAATATTAATTAATGACGCGATATT-----CAATATCTCAAAATAC 69  
QY 41 IleSerGlyGlnTyTyLysProSerValSerValPheSerAsnPheSerValLysGluThr 60  
DB 70 GTTTCATTAATCAATATGGTTTGAACCTTGAATATCCATCAATTCGATAGTAAAGAACAT 129  
QY 61 AsnValIle-----ThrLysAsnLeuIleAlaLeuLys 71  
DB 130 TCAGCAATGTTTGTAGCATCTTTCATGGAATCAAAACGATATAAACTCACCAGTCTTAAT 189  
QY 72 LysAspValAspSerIleGluThrLysThrAspAlaSerValGlyIleSerAsnProSer 91  
DB 190 GTTACAAATTCAAAATTTAGAGAGATCAACTGGATCAGATCAATCTATGACACCAACAA 249  
QY 92 AsnPheThrIleProTyTyThrAlaValPheGlnAspAsnSerValAsnPheAsnGlyThr 111  
DB 250 TTATTAGATATTCAATTCACAAATTCACAAATTAATGCA----- 291  
QY 112 IleGlyTyThrPheAlaGluGlyThrArgValGluIleGluGlySerTyTyGluPhe 131  
DB 292 -----ACCGTATTGAA---ACAGGTTCATGTAATC--- 321  
QY 132 AspValLysAsnProGlyGlyTyTyThrLysSerAspAlaTyArgTyTyPheAlaLeuAla 151  
DB 322 -----GGTTCAACACAGTCACACTTTTTCATCATATATATGCA----- 357  
QY 152 ArgGluMetLysGlyAsnSerPheThrProLysGluLysValSerAsnSerIlePheHis 171  
DB 358 -----CCAGAACAAAAGTAAGAAAT----- 378  
QY 172 ThrValMetArgAsnAspGlyLeuSerIleSerValIleValAsnValCysTyTyAsp 191  
DB 379 -----AAACAATGTTCTTC 393  
QY 192 PheSerLeuAsnAsnLeuSerIleSerProTyTyIleCysGlyGlyAla----- 207  
DB 394 ATAAAGATAAACAATGATATTCATTTTCATATACCTCATCAATATGGTGAATTCACAAAG 453  
QY 208 GlyValAspAlaIleGluPhe-----PheAspValLeuHisIleLys---Phe 222



```

Db      454  ATTTTACAGCATTTGGAAACATTTGTTAAACCTTTTAAAGATTCGAAGCAAAAGGATTT 513
Qy      223  AlatyrlGlnSerLysLeuGlyIleAlaTyrlSerLeuProSerAsnIleSerLeuPheAla 242
Db      514  AAATACACTCAATGGAAGCATTCACATCAAGTATTAAAGCTCATCATCA-----GCA 567
Qy      243  SerLeuTyrlTyrlHisLys-Val-----MetGlyAsnGlnPheLysAsnLeuAs 258
Db      568  GCAACCTATTATTACCAATATTGGGTACCAAGACATGGAATCAAAATCAAAAGAAATCTTAA 627
Qy      258  nValGlnHisValAlaGluLeu 265
Db      628  CATCAAGAAATATACAGACTCTTT 649

RESULT 8
LOCUS   BP001170
DEFINITION BP001170 Nori Satoh unpublished cDNA library, cleavage stage embryo
Ciona intestinalis cDNA clone cici28k04 5', mRNA sequence.
ACCESSION BP001170
VERSION   BP001170.1 GI:19492503
KEYWORDS EST.
SOURCE   Ciona intestinalis.
ORGANISM Ciona intestinalis.
REFERENCE 1 (bases 1 to 728)
AUTHORS Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
TITLE   Expressed genes in Ciona intestinalis
JOURNAL Unpublished (2000)
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
1. 728
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone="cici28k04"
/clone_lib="Nori Satoh unpublished cDNA library, cleavage
stage embryo"
/tissue type="whole animal"
/dev stage="cleavage stage embryo"
/notice="vector: pbluescript SK"

BASE COUNT 208 a 164 c 160 g 196 t
ORIGIN

Alignment Scores:
Pred. No.: 0.176 Length: 728
Score: 96.50 Matches: 55
Percent Similarity: 38.49% Conservative: 47
Best Local Similarity: 20.75% Mismatches: 102
Query Match: 6.45% Indels: 61
DB: 14 Gaps: 10

US-10-062-624-40 (1-293) x BP001170 (1-728)

Qy      5  LeuLysPheThrIleIleAsn---ThrValLeuValCysLeuSerLeuProAsnIle 23
Db      21  CTAAGTTCAATGATGATGAGGTGACGGTTATTGCACTTCTGCTTGTATCGACCAAGCC 80
Qy      24  SerSerLysAlaIleAsnAsnAlaLysLysTyrlTyrlGlyLeuTyrlSerGly 43
Db      81  ATTGAGCGCATCGGTGTACAGCAAGGACCAAAAGGGCCCTTCAATGTGCTGAATTCC 140
Qy      44  GlnTyrlProSerValSerValPheSerAsnPheSerValLys---GluThrAsnVal 62
Db      141  TACCGCGTTCCGGCACAGAGCGCATCAATCAATCAGCATCGAGTGAATTCGTA 200

```

```

Qy      63  IleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLysThrAsp 82
Db      201  GCCAATAAGAACCTTGGAAACGTTTTCATCGG-----TCCGTGTTTACGCTCACTTCT 251
Qy      83  AlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyrlThrAlaValPheGln 102
Db      252  GCCTCGATGGAGTTGGTTAAT-----
Qy      103  AspAsnSerValAsnPheAsnGlyThrIleGlyTyrlThrPheAlaGluGlyThrArgVal 122
Db      273  -----AATGGA-----CATGCATTGCAAGTC 293
Qy      123  GluIleGluGlySerTyrlGluGluPheAspValLys---AsnProGlyGlyTyrlThrLeu 141
Db      294  AACATGGACGGATCTTACACAATATCTCATCGAAGTGTGCTCCCAATGATTATAAGGCT 353
Qy      142  SerAspAlaTyrlArgTyrlPheAlaLeuAlaGluMetLysGlyAsnSerPheThrPro 161
Db      354  GTTCAGTTCCATCTTCACTGGGCTGCTGCCAACAGAGCGAAGGTTCTTGACGATTGGCTT 413
Qy      162  LysGluLysValSerAsnSerIlePheHisThrValMetArgAsn----- 176
Db      414  AACGGCAAGCAATACTTCGCAGAGTTACATGTTCTTTCATTATATATACTAAGTATGCTAGC 473
Qy      177  -----AspGlyLeuSerIleIleSerValIleValAsnVal 188
Db      474  ATTGGGAAGCGGTAAATAAGCCAGATGGCTCGCAGTACTTGGTGTCTTTGTTGATATC 533
Qy      189  CysTyrlAspPheSerLeuAsn---AsnLeuSerIleSerProTyrlIleCysGlyGlyAla 207
Db      534  AATGACGAAACAAACGAGCATTTTAATCTTTTGTCTTAACCCG----- 575
Qy      208  GlyValAspAlaIleGluPhePheAspValLeuHisIleLysPheAlaTyrlGlnSerLys 227
Db      576  -----ATCAACCGTGTCCAATACCAACAGGAAATGACGTATACATCAAG 623
Qy      228  LeuGlyIleAlaTyrlSerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrlHis 247
Db      624  TTTTCGTCGCAAACTTCTTGCCAAACAGACTTGTC-----GAATATTAC 668
Qy      248  LysValMetGlyAsn 252
Db      669  AGATACAGGGGATCA 683

RESULT 9
LOCUS   AL669478
DEFINITION AL669478 directional larval cDNA library Ciona intestinalis cDNA
clone 048ZE07 5', mRNA sequence.
ACCESSION AL669478
VERSION   AL669478
KEYWORDS EST.
SOURCE   Ciona intestinalis.
ORGANISM Ciona intestinalis.
REFERENCE 1 (bases 1 to 837)
AUTHORS Genoscope.
TITLE   Ciona intestinalis directional larval cDNA library
JOURNAL Unpublished (2002)
COMMENT Contact: Genoscope
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
IMPORTANT: this sequence may contain errors. The Ciona intestinalis
library from which the clone was isolated may be contaminated with
cDNAs from bacteria or other Eukarya.
Directional larval cDNA library originate from Dr.M.Branno,
Stazione A.Dohrn, Naples, Italy, and was prepared in
pBluescript2SK+.
Location/Qualifiers
1. 837
/organism="Ciona intestinalis"

```

```

/db_xref="taxon:7719"
/clone="048ZE07"
/clone_lib="directional larval cDNA library"
/notes="Vector: pBluescript2SK+"
BASE COUNT      242 a  188 c  186 g  219 t      2 others
ORIGIN

Alignment Scores:
Pred. No.:      0.302      Length:      837
Score:          95.50      Matches:    55
Percent Similarity: 38.4%      Conservative: 47
Best Local Similarity: 20.75%      Mismatches: 102
Query Match:     6.38%      Indels:    61
DB:              9        Gaps:    10

US-10-062-624-40 (1-293) x AL669078 (1-837)

QY  5 LeuLysPheThrIleLeuAen---ThrValLeuValCysLeuLeuSerLeuProAenIle 23
Db  30 CTAAGTTTCAACATGATGAAGTGAGCGTTATTCAGCTTCTCTGTTATCGACCAAGCC 89
QY  24 SerSerLysAlaIleAenAenAlaLysLysTyrTyrGlyLeuTyrIleSerGly 43
Db  90 ATTGAGCGACATGGGTGTACAGCAAGCAAGGCGCCCTTCCAATTGGCTGAATCC 149
QY  44 GlnTyrLysProSerValSerValPheSerAenPheSerValLys---GluThrAenVal 62
Db  150 TACGCGCTTGGCGCACAGAGCGCAATCACCACATCAGCATGAGCAATTCGGTA 209
QY  63 IleThrLysAenLeuIleAlaLeuLysLysAspValAspSerIleGluThrLysThrAsp 82
Db  210 GCCATAAGAACCTTGGAAAGCTTGTATCGG-----TCCGTGTTACGACATTCCT 260
QY  83 AlaSerValGlyIleSerAenProSerAenPheThrIleProTyrThrAlaValPheGln 102
Db  261 GCCTCGATGGAGTTGGTTAAT----- 281
QY  103 AspAenSerValAenPheAenGlyThrIleGlyTyrThrPheAlaGluGlyThrArgVal 122
Db  282 -----AATGGA-----CATGCATTCGAAGTN 302
QY  123 GluIleGluGlySerTyrGluGluPheAspValLys---AenProGlyGlyTyrThrLeu 141
Db  303 AACATGGACGATCTTACACATATCTGATCGAAGTGTCTCCCAATGATTATAGGCT 362
QY  142 SerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAenSerPheThrPro 161
Db  363 GTTCAGTTCCATCTTCACCTGGGCTGCTGCCAACAGAGCGAAGGTTCTGAGCATTCGCTT 422
QY  162 LysGluLysValSerAenSerIlePheHisThrValMetArgAen----- 176
Db  423 AACGGCAAGCAATACTTCCGACAGATTACATGTTGTTTCATTATAATACTAAGTATGCTAGC 482
QY  177 -----AspGlyLeuSerIleIleSerValIleValAenVal 188
Db  483 ATTGGGAACCGGTGAATAGCCAGATGGCCCTCGCAGTACTTGGTGTCTTTGTGATATC 542
QY  189 CysTyrAspPheSerLeuAen---AenLeuSerIleSerProTyrIleCysGlyGlyAla 207
Db  543 AATGACGAACAACACGACGATTTTAAATCTTTTGCTAAACCCG----- 584
QY  208 GlyValAspAlaIleGluPheAspValLeuHisIleLysPheAlaTyrGlnSerLys 227
Db  585 -----ATCAACCGTGTCCAAATACCAACCCAGGAATGACGTATACATCAAAAG 632
QY  228 LeuGlyIleAlaTyrSerLeuProSerAenIleSerLeuPheAlaSerLeuTyrIleHis 247
Db  633 TTTTCGTCGCAAAATCTTGGCACTGACTGTGCA-----GAATATTAC 677
QY  248 LysValMetGlyAen 252
Db  678 AGATACAGGGGATCA 692

RESULT 10
AL669078 AL669078 862 bp mRNA linear EST 14-JAN-2002
LOCUS AL669078 directional larval cDNA library Ciona intestinalis cDNA
DEFINITION clone 041ZB06 5', mRNA sequence.
ACCESSION AL669078
VERSION AL669078.1 GI:18142335
KEYWORDS EST.
SOURCE Ciona intestinalis.
ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Cionidae; Ciona.
REFERENCE 1 (bases 1 to 862)
AUTHORS Genoscope.
TITLE Ciona intestinalis directional larval cDNA library
JOURNAL Unpublished (2002)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
IMPORTANT: this sequence may contain errors. The Ciona intestinalis
library from which the clone was isolated may be contaminated with
cDNAs from bacteria or other Eukarya.
Directional larval cDNA library originate from Dr M.Branno,
Stazione A.Dohrn, Naples, Italy, and was prepared in
pBluescript2SK+.
FEATURES
Location/Qualifiers
1..862
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone="041ZB06"
/clone_lib="directional larval cDNA library"
/notes="Vector: pBluescript2SK+"
BASE COUNT      242 a  203 c  198 g  216 t      3 others
ORIGIN

Alignment Scores:
Pred. No.:      0.318      Length:      862
Score:          95.50      Matches:    47
Percent Similarity: 37.91%      Conservative: 33
Best Local Similarity: 22.27%      Mismatches: 72
Query Match:     6.38%      Indels:    59
DB:              9        Gaps:    9

US-10-062-624-40 (1-293) x AL669078 (1-862)

QY  84 SerValGlyIleSerAenProSerAenPheThrIleProTyrThrAla----- 99
Db  82 AGCCAAGGACCAAAAGGCGCTTCCAATTGGCTGAATTCCTACGCGCTTGGCGCACAGA 141
QY  100 -----ValPheGlnAspAenSerValAenPheAenGlyThrIleGly 113
Db  142 GCGCAATCACCACATCAGCATCGACATGAAAATTCGGTA---GCCAATAAGAACCTTGA 198
QY  114 TyrThrPheAlaGluGlyThrArgValGluIleGluGlySerTyrGlu----- 129
Db  199 ---ACGTTTGATCGGTCCGTGTTTACGACATCTCTCGCTCGATGAGTTCGTTAATAAT 255
QY  130 -----GluPheAspValLysAenProGlyGlyTyrThrLeuSerAspAla----- 144
Db  256 GGACATGCATTGCAAGTCAACATGGACGATCTTACACAATATCTGATCGAAGTGTGCTC 315
QY  145 -----TyrArgTyrPheAlaLeuAlaArgGluMetLys 155
Db  316 CCCAATGATTATNAGGCTGTTTCAGTTCATCTTCTACTGGCTGTGCCAACAGAGCGAA 375
QY  156 GlyAenSerPheThrProLysGluLysValSerAenSerIlePheHisThrValMetArg 175
Db  376 GGTTCAGCATTCGCTTAACGGCAAGCAATACTTCGACAGACTTACATGTTGTTTATTAT 435
QY  176 Aen-----AspGlyLeuSerIleIle 182
Db  436 AATACTAATTATGCTAGCATTTGGGGAAGCGGTGAATAAGCCAGATGGCTCGCAGTACTT 495

```

```

QY 193 SerValIleValAsnValCysTyrAspPheSerLeuAen---AsnLeuSerIleSerPro 201
Db 496 GGTGCTTTGTATATCAATGACGAAACAAACAGCAGCATTTTAAATCTTTTGTAAACCG 555
QY 202 TyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIleLys 221
Db 556 -----ATCAACCGTGTCCAAATACCAACAGGAA 585
QY 222 PheAlaTyrGlnSerLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPhe 241
Db 586 ATGAGTATACATCAAGATTTTCGTCGCAAAANTCTTGCCAACTGACTTGTC 639
QY 242 AlaSerLeuTyrTyrHisLysValMetGlyAen 252
Db 640 -----GAATATTACAGATACAGGGGATCA 663

RESULT 11
AV938117
LOCUS
DEFINITION AV938117 K. Sato unpublished cDNA library, strain H602 adult,
cDNA clone bahl6d08 5', mRNA sequence.
ACCESSION AV938117
VERSION AV938117.1 GI:18233914
KEYWORDS EST.
ORGANISM Hordeum vulgare subsp. spontaneum.
Eukaryota; Viridiplantae; Hordeum vulgare subsp. spontaneum
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
1 (bases 1 to 648)
Sato, K., Saisho, D. and Takeda, K.
Barley EST sequencing project in NIG and Okayama Univ
Unpublished (2002)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel.: 81-559-81-6856
Fax: 81-559-81-6855
Email: tehinigenes.nig.ac.jp.
FEATURES
Location/Qualifiers
source
1..648
/organism="Hordeum vulgare subsp. spontaneum"
/strain="H602"
/db_xref="taxon:77009"
/clone="bahl6d08"
/clone_lib="K. Sato unpublished cDNA library, strain H602
adult, heading stage top three leaves"
/tissue type="top three leaves"
/dev stages="adult, heading stage"
BASE COUNT 161 a 145 c 169 g 172 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 0.259 Length: 648
Score: 94.50 Matches: 49
Percent Similarity: 36.08% Conservative: 21
Best Local Similarity: 25.26% Mismatches: 99
Query Match: 6.32% Indels: 25
DB: 10 Gaps: 8

US-10-062-624-40 (1-293) x AV938117 (1-648)

QY 40 TyrIleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGlu 59
Db 112 TACATCATCGCGCATTCACGCGCCCTCGCAGCATCTTCGTGACCTTTTCGATCAGAGG 171
QY 60 ThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAsp----- 73
Db 172 -----TGCCGGAACAGGTTTCAGTCAGAGGATATATGGGAACAGCGCGATG 219

```

```

QY 74 -----ValAspSerIleGluThrLysThrAspAlaSerValGlyIleSerAsnPro 90
Db 220 GTGCCGGCGTTCCAGAGCCTGGAGAC---ATAGCCGAGAGGATCATCAATAGCGCTGTT 276
QY 91 SerAsnPheThrIleProTyrThrAlaValPheGlnAspAsnSerValAsnPheAsnGly 110
Db 277 CCTGGTAAAGGATTTGACACATGGGTGTT-----AAGATTGAGCTGCTTGGT 324
QY 111 ThrIleGlyTyrThrPheAlaGluGlyThrArgValGluIleGluGlySerTyrGluGlu 130
Db 325 CAGATAGAGTTGTATTTCGACAGAGCAACTTCTACGACTTCACTTCATTGGTGGTGGT 384
QY 131 PheAspValLysAsnProGlyGlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeu 150
Db 385 TTAGATGTT-----CCTGGTGAATATATGAAAGAAAGACATATCCATTTCAGTTCTCT 438
QY 151 AlaArgGluMetLysGlyAsnSerPheThrProLysGluLysValSerAsnSerIlePhe 170
Db 439 ACTGTTGAATGCGCATATGAGTCATCAATGGAACAATGTCAGACTGAGGTACATCTGT 498
QY 171 HisThrValMetArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysTyr 190
Db 499 AAGTGCAGCAT-----GGTAGAACTATGTTGGCAATATTTGTGAA---TGTCGG 546
QY 191 AspPheSerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAsp 210
Db 547 GATTTCGTGTAGAGCAACTATCTCCGCTTCCTTCGATCAATAACAGCATCAAGATGAA 606
QY 211 AlaIleGluPhePheAspValLeuHisIleLysPheAlaTyr 224
Db 607 ---GTGGATTAAGGATTTGTCGATATTGAGTTTGATGATAC 645

RESULT 12
AV938117
LOCUS
DEFINITION AV938117 K. Sato unpublished cDNA library, strain H602 adult,
cDNA clone bahl6d08 5', mRNA sequence.
ACCESSION AV938117
VERSION AV938117.1 GI:3807690
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 547)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapba-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1067 Std Error: 0.00
Seq Primer: -40UP from Gibco
High quality sequence stop: 414.
Location/Qualifiers
source
1..547
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1858133"
/tissue type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_hosts="DH10B"
/notes="Organ: mixed (see below); Vector: pRTT3D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHW, pregnant uterus
NbHFU, and fetal heart NbHH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization

```

2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA

```

Db      203 TCAGAAATAGAAAGACTTGTATACATATAAGAAATATCGTAATAATCATTAATAGGATTATTA 144
Qy      183 SerValIleValAsnVal 188
Db      143 ATTATGTATATATCTATA 126

RESULT 14
AZ534829/c
LOCUS      880 bp      DNA      linear      GSS 03-NOV-2000
DEFINITION ENTCK35TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION  AZ534829
VERSION     AZ534829.1 GI:11091773
KEYWORDS   GSS.
SOURCE     Entamoeba histolytica.
ORGANISM   Entamoeba histolytica
REFERENCE  1 (bases 1 to 880)
AUTHORS   Loftus B., Van Aken, S. and Fraser, C.
TITLE     Determination of clone end sequences from Entamoeba histolytica
JOURNAL   HMI:IMSS sheared DNA library
COMMENT   Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 26
            High quality sequence stop: 806.
FEATURES   source
            1..880
                /organism="Entamoeba histolytica"
                /strain="HMI:IMSS"
                /db_xref="taxon:5759"
                /clone_lib="Entamoeba histolytica Sheared DNA"
                /note="Vector: pHO51; Site 1: Bst I; Constructed at The
            Institute for Genomic Research (TIGR), Rockville, MD.
            Genomic DNA isolated from broth cultures of E. histolytica
            using a method described by Clark and Diamond (Clark,
            C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
            method for isolate identification. Exp. Parasitol.
            77:450.). The DNA was mechanically sheared to give a
            tight size distribution (~2 kb). The v + i method used for
            the library construction is described in detail in Smith,
            H.O. and Venter, J.C. (Making small insert libraries for
            whole genome shotgun sequencing projects. In Genome
            Sequencing: A Practical Approach, eds M. Vaudin and B.
            Barrell, Oxford University Press, 1999)."
BASE COUNT  290 a 151 c 120 g 319 t
ORIGIN
Alignment Scores:
Pred. No.:      0.802      Length:      880
Score:          92.50      Matches:     27
Percent Similarity: 47.75%      Conservative: 26
Best Local Similarity: 24.32%      Mismatches: 51
Query Match:    6.18%      Indels:      7
DB:             17      Gaps:         4

US-10-062-624-40 (1-293)-x AZ534829 (1-880)

Qy      13 ValLeuValCysLeuLeuSerLeuProAsnIleSerSerSer---LysAlaIleAsnAsn 31
Db      690 GTACTTACCTGTGTTAATAATAAACAAGTAATACTGTGTAATTTAAATTCGTGTACAAAT 631
Qy      32 AsnAlaIleLysTyrTyrTyrGlyLeuTyrIleSerGlyGlnTyrLysProSerValSerVal 51

```

```

Db      630 AACGGTAAAGACAGACAGAGTTATGGTATTCGAATTAATAATATCCAGACGAAACTCA 571
Qy      52 PheSerAsnPheSerValLysGluThrAsnValIleThrLysAsnLeuIleAlaLeuLys 71
Db      570 ACTATAAATGCTAATTTCAAGCATAAATGGAGTAATCGTTCTTAATGTGTTGGAAATGGA 511
Qy      72 LysAspVal-----AspSerIleGluThrLysThrAspAlaSerValGlyIleSerAsn 89
Db      510 GGTAATGTTAAAAACGATAAACTAGATCTTAGTCTCTAATGCTGAGTTTGGAAAGAATAAC 451
Qy      90 --ProSerAsnPheThrIleProTyrThrAlaValPheGlnAspAsnSerValAsnPhe 108
Db      450 AAAAAGCGAGACCTTACCTTTCAATAAT-----CCATCTGATAAATCTCTGTT 400
Qy      109 AsnGlyThrIleGlyTyrThrPheAlaGluGly 119
Db      399 GATGGTAAAGTTGGTTATCTTATGCTAATGGA 367

RESULT 15
BJ426331
LOCUS      584 bp      mRNA      linear      EST 11-MAR-2002
DEFINITION BJ426331 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv58j17 5', mRNA sequence.
ACCESSION  BJ426331
VERSION     BJ426331.1 GI:19343038
KEYWORDS   EST.
SOURCE     Dictyostelium discoideum.
ORGANISM   Dictyostelium discoideum
REFERENCE  1 (bases 1 to 584)
AUTHORS   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
TITLE     Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
            Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
JOURNAL    Unpublished (2002)
COMMENT    Contact: Tadaasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tahini@genes.nig.ac.jp.
FEATURES   source
            1..584
                /organism="Dictyostelium discoideum"
                /strain="AX4"
                /db_xref="taxon:44689"
                /clone="ddv58j17"
                /clone_lib="Dictyostelium discoideum cDNA library, VF"
                /sex="mat A"
                /dev_stage="Growth phase"
BASE COUNT  213 a 220 g 220 t 1 others
ORIGIN
Alignment Scores:
Pred. No.:      0.454      Length:      584
Score:          92.00      Matches:     51
Percent Similarity: 37.00%      Conservative: 33
Best Local Similarity: 22.47%      Mismatches: 83
Query Match:    6.15%      Indels:      60
DB:             13      Gaps:         11

US-10-062-624-40 (1-293) x BJ426331 (1-584)

Qy      9 IleIleAsnThrValLeuValCysLeu-----LeuSerLeuProAsnIleSer 24
Db      19 GTNATCAATATTGGTATATGTAATTAATTAATTCGTATATAGTATATATATATATATCT 78
Qy      25 SerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyrIleSerGlyGln 44
Db      79 ATTTTAAACAGATCAACAATAAGAAAAA-----ATGAAAAGATCA 120
Qy      45 TyrLysProSerValSerValPheSerAsnPheSerValLysGluThrAsnValIleThr 64

```

```

Db      121 GAAAAATCAAGTACATCTGTTAGTAATAACAAACAAAGATGTAATAATATCATCAGT 180
Qy      65 LysAsnLeuAlaLeuLysLysValAspSerIleGluThrLysThrAspAlaSer 84
Db      181 TCAAAATGAAGTTGGTGTAAAGAA-----GAAAAATAA----- 213
Qy      85 ValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaValPheGlnAspAsn 104
Db      214 -----GGACATCAAGAAATCTTATTAAAGTTTTTAATCTTA----- 249
Qy      105 SerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArg----- 121
Db      250 -----TCAGTCATTATGTTTGGTATTTAGCATTTCCTCACTGTTTATTCTCA 291
Qy      122 ValGluIleGluGlySerTyrGluGluPheApp----- 132
Db      292 GTATTACGTTATCAAGAGTGTATTATCATGAATTTGATCCCATATTTTAATTTATAGATCAACA 351
Qy      133 -----ValLysAsnProGlyGlyTyrThrLeu-----SerAspAlaTyrArgTyr 147
Db      352 ATATATCTTTGTCAGAGAGTTTTTATAATTTTAAATGTTGATGAAAGAGCATGG 411
Qy      148 PheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGluLysValSerAsn 167
Db      412 TATCCATTAGGACGT---ATTGTAGGTGTGTACAAATTTACCAGGTTTAAATGCGCAACAGCA 468
Qy      168 SerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSerValIleValAsn 187
Db      469 AGTTTAGTTCATTGGTGCATTG-----AATTCATTGAATATTACAGTTTAATATTAGAAAT 522
Qy      188 ValCysTyrAspPheSerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAla 207
Db      523 GTATGT-----GTATGTTATCACCATTGGTTTCATCAAAATACA 561
Qy      208 GlyValAspAlaIleGluPhe 214
Db      562 GCAATGGTAACCTATAAATTT 582

RESULT 16
BJ420538 594 bp mRNA linear EST 10-MAR-2002
LOCUS BJ420538 Dictyostelium discoideum cDNA library, VF Dictyostelium
DEFINITION dictyostelium cDNA clone d4v39j13 5', mRNA sequence.
ACCESSION BJ420538
VERSION BJ420538.1 GI:19332778
KEYWORDS Dictyostelium discoideum.
SOURCE Dictyostelium discoideum
ORGANISM Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 594)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the vegetative
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadao Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
Location/Qualifiers
1..594
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="d4v39j13"
/clone_lib="Dictyostelium discoideum cDNA library, VF"
/sex="mat A"
/dev_stage="Growth phase"
BASE COUNT 217 a 62 c 90 g 225 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 0.467 Length: 594
Score: 32.00 Matches: 51
Percent Similarity: 37.00% Conservative: 33
Best Local Similarity: 22.47% Mismatches: 83
Query Match: 6.15% Indels: 60
DB: 13 Gaps: 11

US-10-062-624-40 (1-293) x BJ420538 (1-594)
Qy      9 IleIleAsnThrValLeuValCysLeu-----LeuSerLeuProAsnIleSer 24
Db      26 GTCATCAATATGTTGTGTATATGATAATAATTTGTTATATTTAGTGTATATATATATCT 85
Qy      25 SerSerLysAlaIleAsnAsnAsnAlaLysLysTyrTyrGlyLeuTyrIleSerGlyGln 44
Db      86 ATTTTAAACAGATCAACAATTAAGAAAAA-----ATGAAAAGATCA 127
Qy      45 TyrLysProSerValSerValPheSerAsnPheSerValLysGluThrAsnValIleThr 64
Db      128 GAAAAATCAAGTACATCTGTTGTTAGTAATAACAAACAAAGATGTAATAATATCATCAGT 187
Qy      65 LysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLysThrAspAlaSer 84
Db      188 TCAAAATGAAGTTGGTGTAAAGAA-----GAAAAATAA----- 220
Qy      85 ValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaValPheGlnAspAsn 104
Db      221 -----GGACATCAAGAAATCTTATTAAAGTTTTTAATCTTA----- 256
Qy      105 SerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArg----- 121
Db      257 -----TCAGTCATTATGTTTGGTATTTAGCATTTCCTCACTGTTTATTCTCA 298
Qy      122 ValGluIleGluGlySerTyrGluGluPheApp----- 132
Db      299 GTATTACGTTATCAAGAGTGTATTATCATGAATTTGATCCCATATTTTAATTTATAGATCAACA 358
Qy      133 -----ValLysAsnProGlyGlyTyrThrLeu-----SerAspAlaTyrArgTyr 147
Db      359 ATATATCTTTGTCAGAGAGTGTATTATAATTTTAAATGTTGATGAAAGAGCATGG 418
Qy      148 PheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGluLysValSerAsn 167
Db      419 TATCCATTAGGACGT---ATTGTAGGTGTGTACAAATTTACCAGGTTTAAATGCGCAACAGCA 475
Qy      168 SerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSerValIleValAsn 187
Db      476 AGTTTAGTTCATTGGTGCATTG-----AATTCATTGAATATTACAGTTTAATATTAGAAAT 529
Qy      188 ValCysTyrAspPheSerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAla 207
Db      530 GTATGT-----GTATGTTATCACCATTGGTTTCATCAAAATACA 568
Qy      208 GlyValAspAlaIleGluPhe 214
Db      569 GCAATGGTAACCTATAAATTT 589

RESULT 17
BJ336589 595 bp mRNA linear EST 05-MAR-2002
LOCUS BJ336589 Dictyostelium discoideum cDNA library, AF Dictyostelium
DEFINITION dictyostelium cDNA clone dda54n07 5', mRNA sequence.
ACCESSION BJ336589
VERSION BJ336589.1 GI:19166719
KEYWORDS Dictyostelium discoideum.
SOURCE Dictyostelium discoideum
ORGANISM Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 595)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the aggregation
stage

```

[illegible]

```

Qy 122 ---ValGluLeuGluGlySerTyrGluGluPheAsp----- 132
Db 281 GTATTACGTTATGAAAGTGTATTATCATGAATTTGATCATATTTTAAATTTATAGATCAACA 340
Qy 133 -----ValLysAsnProGlyGlyTyrThrLeu-----SerAspAlaTyrArgTyr 147
Db 341 ATATATCTTGTTCAGAGAGTTTATATATATTTTAAATTTGATGTTGATGAAAGCATGG 400
Qy 148 PheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGluLysValSerAsn 167
Db 401 TATCCATTAGGACGT--ATTGTAGTGTGTACAAATTTACCCAGGTTTAAATGCAACACGCA 457
Qy 168 SerIlePheHisThrValMetArgAsnAepGlyLeuSerIleIleSerValIleValAsn 187
Db 458 AGTTAGTTTCAATGTCATG--AAATTCATTGAATATATACAGTTTAAATATAGAAAT 511
Qy 188 ValCysTyrAspPheSerLeuAsnLeuSerIleSerProTyrIleCysGlyGlyAla 207
Db 512 GTATGT-----GTATTGTTATCACCATGTTGTCATCAAAATACA 550
Qy 208 GlyValAspAlaIleGluPhe 214
Db 551 GCAATGGTAACCTATAAATTT 571

RESULT 19
BJ426166 625 bp mRNA linear EST 11-MAR-2002
LOCUS BJ426166 Dictyostelium discoideum cDNA library, VF Dictyostelium
DEFINITION BJ426166 Dictyostelium discoideum cDNA clone d4v5j04 5', mRNA sequence.
ACCESSION BJ426166
VERSION BJ426166.1 GI:19342873
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE 1 (bases 1 to 625)
AUTHORS Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
TITLES Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
Full length cDNA of Dictyostelium discoideum at the vegetative
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6855
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
FEATURES
source
1..625
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="d4v58j04"
/clone_lib="Dictyostelium discoideum cDNA library, VF"
/sex="mat A"
/dev_stage="growth phase"
BASE COUNT 228 a 70 c 102 g 225 t
ORIGIN

Alignment Scores:
Pred. No.: 0.511 Length: 625
Score: 92.00 Matches: 51
Percent Similarity: 37.00% Conservative: 33
Best Local Similarity: 22.47% Mismatches: 83
Query Match: 6.15% Indels: 60
DB: 13 Gaps: 11

US-10-062-624-40 (1-293) x BJ426166 (1-625)

Qy 9 lleIleAsnThrValLeuValCysLeu-----LeuSerLeuProAsnIleSer 24
Db 2 GTCATCAATATTGTTGTATATGTTATATTAATTTGTTATATTAGTTGTTATATATATCT 61

```

```

Qy 25 SerSerLysAlaIleAsnAsnAsnAlaLysLysTyrTyrGlyLeuTyrIleSerGlyGln 44
Db 62 ATTTTAAACAGATCAACAATAAGAAAAA-----ATGAAAAGATCA 103
Qy 45 TyrLysProSerValSerValPheSerAsnPheSerValLysGluThrAsnValIleThr 64
Db 104 GAAAAATCAAGTACATCTGTTGTAGTAACAACAACAGATGTAATATATCATCAGT 163
Qy 65 LysAsnLeuIleAlaLeuLysLysAspSerIleGluThrLysThrAspAlaSer 84
Db 164 TCAAAATCAAGTGTGTTAAAGAA-----GAAAAATAAA----- 196
Qy 85 ValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaValPheGluAspAsn 104
Db 197 -----GGACATCAAGAAATCTTATTAAAGGTTTAAATTTCTA----- 232
Qy 105 SerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArg----- 121
Db 233 -----TCAGTCATTATGTTTGTAGCATTTTCACTCGTTTATTCTCA 274
Qy 122 ---ValGluLeuGluGlySerTyrGluGluPheAsp-----SerAspAlaTyrArgTyr 147
Db 275 GTATTACGTTATGAAAGTGTATTATCATGAATTTGATCATATTTTAAATTTATAGATCAACA 334
Qy 133 -----ValLysAsnProGlyGlyTyrThrLeu-----SerAspAlaTyrArgTyr 147
Db 335 ATATATCTTGTTCAGAGAGTTTATATAATTTTAAATTTGATGTTGATGAAAGCATGG 394
Qy 148 PheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGluLysValSerAsn 167
Db 395 TATCCATTAGGACGT--ATTGTAGTGTGTACATTTACCAGGTTTAAATGCAACACGCA 451
Qy 168 SerIlePheHisThrValMetArgAsnAepGlyLeuSerIleIleSerValIleValAsn 187
Db 452 AGTTAGTTTCAATGTCATG--AAATTCATTGAATATATACAGTTTAAATATAGAAAT 505
Qy 188 ValCysTyrAspPheSerLeuAsnLeuSerIleSerProTyrIleCysGlyGlyAla 207
Db 506 GTATGT-----GTATTGTTATCACCATGTTGTCATCAAAATACA 544
Qy 208 GlyValAspAlaIleGluPhe 214
Db 545 GCAATGGTAACCTATAAATTT 565

RESULT 20
BJ424116 641 bp mRNA linear EST 11-MAR-2002
LOCUS BJ424116 Dictyostelium discoideum cDNA library, VF Dictyostelium
DEFINITION BJ424116 Dictyostelium discoideum cDNA clone d4v5ln07 5', mRNA sequence.
ACCESSION BJ424116
VERSION BJ424116.1 GI:19340823
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE 1 (bases 1 to 641)
AUTHORS Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
TITLES Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
Full length cDNA of Dictyostelium discoideum at the vegetative
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
FEATURES
source
1..641
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"

```



```
/clone="ddv5ln07"
/clone_lib="Dictyostelium discoideum cdna library, VF"
/sex="mat A"
/dev_stage="Growth phase"
BASE COUNT 232 a 70 c 104 g 235 t
ORIGIN

Alignment Scores:
Pred. No.: 0.534 Length: 641
Score: 92.00 Matches: 51
Percent Similarity: 37.00% Conservativeness: 33
Best Local Similarity: 22.47% Mismatches: 83
Query Match: 6.15% Indels: 60
DB: 13 Gaps: 11

US-10-062-624-40 (1-293) x BJ424116 (1-641)

Qy 9 IleIleAenThrValLeuValCysLeu-----LeuSerLeuProAsnIleSer 24
Db 19 GTCATCAATATTGTTGTATATGTAATAATAATTGTTATATTAGTGTATAATATATATCT 78
Qy 25 SerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyrIleSerGlyGln 44
Db 79 ATTTTAAACAGATCAACAATAAGAAAA-----ATGAAAGATCA 120
Qy 45 TyrLysProSerValPheSerAsnPheSerValLysGluThrAsnValIleThr 64
Db 121 GAAAAATCAAGTACATCTGTTAGTAAATAACAAACAAGATGTAATATCATCATCT 180
Qy 65 LysAsnLeuIleAlaLeuLysLysValAspSerIleGluThrLysThrAspAlaSer 84
Db 181 TCAATGAAGTGGTGTAAAGAA-----GAAAAATAA- 213
Qy 85 ValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaValPheGlnAspAsn 104
Db 214 -----GNCATCAGAAATCTTATAAAGTTTAAATCTTA----- 249
Qy 105 SerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArg- 121
Db 250 -----TCAGTCATTATGTTTATGATCATGATTTTCAACTCGTTTATCTCA 291
Qy 122 ---ValGluIleGluGlySerTyrGluGluPheAsp----- 132
Db 292 GTATTACGTTATGAAAGGTTTATTCATGAAATTTGATCATCATATTTTAAATATATAGATCAACA 351
Qy 133 -----ValLysAsnProGlyGlyTyrThrLeu-----SerAspAlaTyrArgTyr 147
Db 352 ATATATCTTGTTCAGAGGTTTATTAATTTTAAATTTTAAATTTGTTGATGAAAGGATGG 411
Qy 148 PheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGluLysValSerAsn 167
Db 412 TATCATTTAGGACGT---ATTGAGTGGTACAAATTTACCAGGTTTAAATGCGAACACGCA 468
Qy 168 SerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSerValIleValAsn 187
Db 469 AGTTAGTTCATCGTCAATG-----AATTCATTTGAATATATACAGTTTAAATATAGAAAT 522
Qy 188 ValCysTyrAspPheSerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAla 207
Db 523 GTATGT-----GTATTTGATCACCAGTGGTTGTCATCAAAATACA 561
Qy 208 GlyValAspAlaIleGluPhe 214
Db 562 GCAATGGTGAACCTATATAATTT 582

RESULT 21
BJ420225 LOCUS
DEFINITION BJ420225 Dictyostelium discoideum cdna library, VF Dictyostelium
discoideum cdna clone ddv38j17 5', mRNA sequence.
ACCESSION BJ420225
VERSION BJ420225.1 GI:19332465
KEYWORDS EST.
```

```
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
REFERENCE Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
1 (bases 1 to 644)
AUTHORS Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
TITLE Full length cDNA of Dictyostelium discoideum at the vegetative
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tehinig@genes.nig.ac.jp.
FEATURES
source
1..644
Location/Qualifiers
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddv38j17"
/clone_lib="Dictyostelium discoideum cdna library, VF"
/sex="mat A"
/dev_stage="Growth phase"
BASE COUNT 234 a 71 c 103 g 236 t
ORIGIN

Alignment Scores:
Pred. No.: 0.538 Length: 644
Score: 92.00 Matches: 51
Percent Similarity: 37.00% Conservativeness: 33
Best Local Similarity: 22.47% Mismatches: 83
Query Match: 6.15% Indels: 60
DB: 13 Gaps: 11

US-10-062-624-40 (1-293) x BJ420225 (1-644)

Qy 9 IleIleAenThrValLeuValCysLeu-----LeuSerLeuProAsnIleSer 24
Db 26 GTCATCAATATTGTTGTATATGTAATAATAATTGTTATATTAGTGTATAATATATATCT 85
Qy 25 SerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyrIleSerGlyGln 44
Db 86 ATTTTAAACAGATCAACAATAAGAAAA-----ATGAAAGATCA 127
Qy 45 TyrLysProSerValSerValPheSerAsnPheSerValLysGluThrAsnValIleThr 64
Db 128 GAAAAATCAAGTACATCTGTTAGTAAATAACAAACAAGATGTAATATCATCATCT 187
Qy 65 LysAsnLeuIleAlaLeuLysLysValAspSerIleGluThrLysThrAspAlaSer 84
Db 188 TCAATGAAGTGGTGTAAAGAA-----GAAAAATAA- 220
Qy 85 ValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaValPheGlnAspAsn 104
Db 221 -----GNCATCAGAAATCTTATAAAGTTTAAATCTTA----- 256
Qy 105 SerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArg- 121
Db 257 -----TCAGTCATTATGTTTATGATCATGATTTTCAACTCGTTTATCTCA 298
Qy 122 ---ValGluIleGluGlySerTyrGluGluPheAsp----- 132
Db 299 GTATTACGTTATGAAAGGTTTATTCATGAAATTTGATCATCATATTTTAAATATATAGATCAACA 358
Qy 133 -----ValLysAsnProGlyGlyTyrThrLeu-----SerAspAlaTyrArgTyr 147
Db 359 ATATATCTTGTTCAGAGGTTTATTAATTTTAAATTTTAAATTTGTTGATGAAAGGATGG 418
Qy 148 PheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGluLysValSerAsn 167
Db 419 TATCCATTAGGACGT---ATTGAGTGGTGGCACAATTTACCAGGTTTAAATGCGAACACGCA 475
```

```

Qy 168 SerIlePheHisThrValMetArgAsnAspGlyLeuSerIleSerValIleValAsn 187
Db 476 AGTTTAGTTCATGTCATG-----AATTCATTGAATATTACAGTTAAATATTAGAAAT 529
Qy 188 ValCysTyrAspPheSerLeuAsnLeuSerIleSerProTyrIleCysGlyGlyAla 207
Db 530 GTATGCT-----GTATTGTTATCACCATGTTGTCATCAATACA 568
Qy 208 GlyValAspAlaIleGluPhe 214
Db 569 GCAATGGTAACCTATAAATTT 589

RESULT 22
LOCUS BJ418961
DEFINITION BJ418961 Dictyostelium discoideum cDNA library, VF Dictyostelium
ACCESSION BJ418961
VERSION BJ418961
KEYWORDS BJ418961.1 GI:19331201
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 645)
AUTHORS Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
TITLE Full length cDNA of Dictyostelium discoideum at the vegetative
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
source
1..645
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddv34114"
/sex="mat A"
/dev_stage="Growth phase"
BASE COUNT 232 a 70 c 104 g 239 t
ORIGIN

Alignment Scores:
Pred. No.: 0.54 Length: 645
Score: 92.00 Matches: 51
Percent Similarity: 37.00% Conservative: 33
Best Local Similarity: 22.47% Mismatches: 83
Query Match: 6.15% Indels: 60
DB: 13 Gaps: 11

US-10-062-624-40 (1-293) x BJ418961 (1-645)
Qy 9 IleIleAsnThrValLeuValCysLeu-----LeuSerLeuProAsnIleSer 24
Db 22 GTCATCAATATTGTTGTTATGTTATATAAATTTGTTATATTAGTTATATATATATCT 81
Qy 25 SerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyrIleSerGlyGln 44
Db 82 ATTTTAAACACATCAACATTAAGAAAAA-----ATGAAAGATCA 123
Qy 45 TyrLysProSerValSerValPheSerAsnPheSerValLysGluThrAsnValIleThr 64
Db 124 GAAATAATCAAGTACATCTGTTGTTAGTATACAAACACACAGATGTAATATCATCAT 183
Qy 65 LysAsnLeuIleAlaLeuLysLysValAspSerIleGluThrLysThrAspAlaSer 84
Db 184 TCAAAATGAAGTTGGTTAAAGAA-----GAAATATAA----- 216

```

```

Qy 85 ValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaValPheGlnAspAsn 104
Db 217 -----GGACATCAAGAATCTTATTTAAAGTTTAAATCTTA----- 252
Qy 105 SerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArg----- 121
Db 253 -----TCAGTCATTTATGTTTATAGCATTTTCAACATTCGTTATTCTCA 294
Qy 122 ---ValGluIleGluGlySerTyrGluGluPheAsp----- 132
Db 295 GTATTAGCTTATGAAGTGTATTATCATCAATTTGATCCATATTTTAATTATAGATCAACA 354
Qy 133 -----ValLysAsnProGlyGlyTyrThrLeu-----SerAspAlaTyrArgTyr 147
Db 355 ATATATCTTGTTCAGAGGTTTTTATAATTTTAAATGTTTGTATGAAGAGCATGG 414
Qy 148 PheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGluLysValSerAsn 167
Db 415 TATCCATTAGACCT---ATTGTAGTGGTGTACATTTACCACAGGTTTAATGCGAACAGCA 471
Qy 168 SerIlePheHisThrValMetArgAsnAspGlyLeuSerIleSerValIleValAsn 187
Db 472 AGTTTAGTTCATGTCATG-----AATTCATTGAATATTACAGTTAAATATTAGAAAT 525
Qy 188 ValCysTyrAspPheSerLeuAsnLeuSerIleSerProTyrIleCysGlyGlyAla 207
Db 526 GTATGT-----GTATTGTTATCACCATGTTGTCATCAATAACA 564
Qy 208 GlyValAspAlaIleGluPhe 214
Db 565 GCAATGGTAACCTATAAATTT 585

RESULT 23
LOCUS BG594518
DEFINITION EST493196 cSTS Solanum tuberosum cDNA clone cSTS7P13 5' sequence,
ACCESSION BG594518
VERSION BG594518.1 GI:13612658
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 702)
AUTHORS van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Chiemingo, A.,
Bouqri, O., Buell, C.R., Ronning, C., Tanksley, S. and Baker, B.
TITLE Generations of ESTs from sprouting potato eyes
JOURNAL Unpublished (2000)
COMMENT Contact: Cathy Ronning
The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13P-R.
FEATURES
Location/Qualifiers
1..702
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cSTS7P13"
/clone_lib="cSTS"
/tissue_type="sprouting eyes from tubers"
/dev_stage="12-14 weeks post harvest"
/lab_host="SOLR"
/notes="Vector: pBluescript-SK(-); Site 1: EcoRI; Site 2:
XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
taken from tubers. The tubers were incubated at 26C in the
dark for 2-3 weeks prior to sprouting. The eyes were
frozen in liquid nitrogen immediately upon removal from
tubers."
BASE COUNT 222 a 151 c 150 g 179 t
ORIGIN

```



371	ATAAGAATAACAAATGTATTTCATCATCTCATCAAAATGGTGAATTCACAAG	371
208	GlyValAspAlaIleGluPhe-----PheAspValLeuHisIleLys---Phe	222
431	AAATTACCACATTTGGAAACATTTGTTTAAACCTTTTAAAGATTTTGAAGCAAAAGGATTT	490
223	AlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAla	242
491	AAATACACTCAAATGGAAGCATTACATCAAGTAGTATTAAAGCTCATCATCA-----GCA	544
243	SerLeuTyrTyrHisLys-Val-----MetGlyAsnGlnPheLysAsnLeuAs	258
545	GCAACCTATTATTACCAATATTGGGTACCAAGACATGGAATCAAAAAAAGAATCTTAA	604
258	nValGlnHisValAlaGluLeu 265	
605	CATCAAGAATATACAGACTCTT 626	
RESULT 25		
BJ410964		
LOCUS	649 bp mRNA linear EST 10-MAR-2002	
DEFINITION	BJ410964 Dictyostelium discoideum cDNA library, VF Dictyostelium	
ACCESSION	dictoideum cDNA clone ddv2b18 3', mRNA sequence.	
VERSION	BJ410964	
KEYWORDS		
SOURCE	BJ410964.1 GI:19323659	
ORGANISM	EST.	
REFERENCE	Dictyostelium discoideum.	
AUTHORS	Dictyostelium discoideum.	
TITLE	Eukaryota; Mycetozoa; Dictyostelid; Dictyostelium.	
JOURNAL	1 (bases 1 to 649)	
COMMENT	Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.	
FEATURES	Full length cDNA of Dictyostelium discoideum at the vegetative	
source	stage	
	Unpublished (2002)	
	Contact: Tadasu Shin-i	
	Center For Genetic Resource Information	
	National Institute of Genetics	
	1111 Yata, Mishima, Shizuoka 411-8540, Japan	
	Tel: 81-559-81-6856	
	Fax: 81-559-81-6855	
	Email: tshini@gene.nig.ac.jp.	
	Location/Qualifiers	
	1. 649	
	/organism="Dictyostelium discoideum"	
	/strain="AX4"	
	/db_xref="taxon:44689"	
	/clone="ddv2b18"	
	/clone_lib="Dictyostelium discoideum cDNA library, VF"	
	/sex="mat A"	
	/dev stage="Growth phase"	
BASE COUNT	280 a 110 c 76 g 182 t 1 others	
ORIGIN		
Alignment Scores:		
Pred. No.:	0.633	Length: 649
Score:	91.50	Matches: 59
Percent Similarity:	34.70%	Conservative: 34
Best Local Similarity:	22.01%	Mismatches: 91
Query Match:	6.12%	Indels: 84
DB:..	13	Gaps: 12
US-10-062-624-40 (1-293) x BJ410964 (1-649)		
Qy	21 ProAsnIleSerSerLysAlaIleAsnAsnAlaLysTyrTyrGlyLeuTyr 40	
Db	18 CCAAAATATTAAATGAGCGATT-----CAATACTCAAAATAC 56	
Qy	41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60	
Db	57 GTTTCAAATCAATATGTTTGAACCTTCAATATCCATCAATTTGGATAGTAAAGGACAT 116	
Qy	61 AsnValIle-----ThrLysAsnLeuIleAlaLeuLys 71	

```

/db_xref="taxon:44689"
/clone_lib="Dictyostelium discoideum cDNA library, VP"
/sex="mat A"
/dev_stage="Growth phase".
BASE COUNT 275 a 110 c 79 g 186 t
ORIGIN

Alignment Scores:
Pred. No.: 0.635 Length: 650
Score: 91.50 Matches: 59
Percent Similarity: 34.70% Conservative: 34
Best Local Similarity: 22.01% Mismatches: 91
Query Match: 6.12% Indels: 84
DB: 13 Gaps: 12

US-10-062-624-40 (1-293) x BU412332 (1-650)

QY 21 ProAsnIleSerSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40
DB 8 CCAAAATATTAAATGAGCGATATT-----CAATACTCAAAATAC 46
QY 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60
DB 47 GTTCAATCAATATGTTTGAACCTTGAATTCATCAATGATGATAGTAAGGAACAT 106
QY 61 AsnValIle-----ThrLysAsnLeuIleAlaLeuLys 71
DB 107 TCAGCAATGTTTATGACATCTTCATGGAATCAACGATAAATACTACCAAGTCTTAAT 166
QY 72 LysAspValAspSerIleGluThrLysThrAspAlaSerValGlyIleSerAsnProSer 91
DB 167 GTTCAATCAATCAATTTAGAGGATCACTGGATCAGATCAATCTATGACACCAACAA 226
QY 92 AsnPheThrIleProTyrThrAlaValPheGlnAspAsnSerValAsnPheAsnGlyThr 111
DB 227 TTATTAGATATTTCAATTCACAAATCAACAAATTAATGCA----- 268
QY 112 IleGlyTyrThrPheAlaGluGlyThrArgValGluIleGluGlySerTyrGluGluPhe 131
DB 269 -----ACCATATTGAA---ACAGGTTCAATGTAATC--- 298
QY 132 AspValLysAsnProGlyGlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAla 151
DB 299 -----GGTTCAACAGTCAGACTTTTATCATATATGCA----- 334
QY 152 ArgGluMetLysGlyAsnSerPheThrProLysGluLysValSerAsnSerIlePheHis 171
DB 335 -----CCGAACAAAGTAAGAAAT----- 355
QY 172 ThrValMetArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysTyrAsp 191
DB 356 -----AAACATGTTTCTTC 370
QY 192 PheSerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAla----- 207
DB 371 ATAAAGAAATAACAATGATTCATTCATCATACCTCATCAATGGTGAATTCACAAAG 430
QY 208 GlyValAspAlaIleGluPhe-----PheAspValLeuHisIleLys---Phe 222
DB 431 AATTTCACAGCATGGGAACATGTTGTAACACCTTTAAGAAATTCGAAGCAAAAGGATTT 490
QY 223 AlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAla 242
DB 491 AAATACACTCAATGGAAGCATTCATCAATGATTAATAAGCTCATCATCA-----GCA 544
QY 243 SerLeuTyrTyrHisLys-Val-----MetGlyAsnGlnPheLysAsnLeuAs 258
DB 545 GCAACCTATTATTACCAATATTGGTACCAAGACATGGAAATCAAAAAAGAAATCTAAA 604
QY 258 nValGlnHisValAlaGluLeu 265
DB 605 CATCAAGAAATATACAGACTCTT 626

```

## RESULT 27

AL668855

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 918

/organism="Ciona intestinalis"

/db\_xref="taxon:7719"

/clone="0382B04"

/clone\_lib="directional larval cDNA library"

/notes="vector: pBluescript2SK+"

BASE COUNT 259 a 216 c 196 g 236 t 11 others

ORIGIN

Alignment Scores:

Pred. No.: 1.16 Length: 918

Score: 91.50 Matches: 54

Percent Similarity: 37.74% Conservative: 46

Best Local Similarity: 20.38% Mismatches: 104

Query Match: 6.12% Indels: 61

DB: 9 Gaps: 10

US-10-062-624-40 (1-293) x AL668855 (1-918)

QY 5 LeuLysPheThrIleIleAsn---ThrValLeuValCysLeuLeuSerLeuProAsnIle 23

DB 31 CTAAGTTTCAATGATGAAGGTGACGGTTATTGCACTTCTGCTTGTATCGACCAAGCC 90

QY 24 SerSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyrIleSerGly 43

DB 91 ATGGAGCGCATGGGTGTACAGCAAGCAAGGCGCTTCCAATTCGCTGAATTC 150

QY 44 GlnTyrLysProSerValSerValPheSerAsnPheSerValLys---GluThrAsnVal 62

DB 151 TACNCCCTTGCNGCACACAGAGCCCAATCAATCAGCATCGAGCATGAATTCGGTA 210

QY 63 IleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLysThrAsp 82

DB 211 GCCAATAAGAACCTTGGACGTTTGTATCGG-----TCCGTTTGTACGTCACCTCCT 261

QY 83 AlaSerValGlyLysSerAsnProSerAsnPheThrIleProTyrThrAlaValPheGln 102

DB 262 GCCTCGATGAGTGGTGAAT----- 282

QY 103 AspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArgVal 122

DB 283 -----AATGGA-----CATGCATTGCAAGTC 303

QY 123 GluIleGluGlySerTyrGluGluPheAspValLys---AsnProGlyGlyTyrThrLeu 141

Db 304 AACATGGACGGATCTTACACATATCTGATCGAAGTGTGCCCCCAATGATTATNAGGCT 363  
 Qy 142 SerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrPro 161  
 Db 364 GTNCAAGTTCATCTTCACTGGGCTGCTGCCAACAAGAGCCAGGTTCTGAGCNTGGCTT 423  
 Qy 162 LysGluLysValSerAsnSerIlePheHisThrValMetArgAsn----- 176  
 Db 424 AACGCCAAGCNAATCTTCGACAGATTACATGTTGTTCAATTAATTAATTAAGTATGCCAGC 483  
 Qy 177 -----AspGlyLeuSerIleSerValIleValAsnVal 188  
 Db 484 ATGGGGAGCGGTAAATAAGCCANANGCCCTCCAGTCTGCTGTTGTTGATATC 543  
 Qy 189 CysTyrAspPheSerLeuAsn---AsnLeuSerIleSerProTyrIleCysGlyGlyAla 207  
 Db 544 AATGACGAATCAACAGCAGCATTTTAATCTTTGCTAAACCCG----- 585  
 Qy 208 GlyValAspAlaIleGluPhePheAspValLeuHisIleLysPheAlaTyrClnSerLys 227  
 Db 586 -----ATCAACCGTGTCCAAATACCAACAGGAATGACGTATACATCAAG 633  
 Qy 228 LeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrHis 247  
 Db 634 TTTTCCGTGCGAAATCTTGCCACAGACTGTGCA-----GAATATTAC 678  
 Qy 248 LysValMetGlyAsn 252  
 Db 679 AGATACAGGGGATCA 693

RESULT 28  
 BJ423418 584 bp mRNA linear EST 11-MAR-2002  
 LOCUS BJ423418 Dictyostelium discoideum cDNA library, VF Dictyostelium  
 DEFINITION discoideum cDNA clone ddv49k01 5', mRNA sequence.

ACCESSION BJ423418  
 VERSION BJ423418  
 KEYWORDS  
 ORGANISM Dictyostelium discoideum.  
 REFERENCE 1 (bases 1 to 584)  
 AUTHORS Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.  
 TITLES Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.  
 Full length cDNA of Dictyostelium discoideum at the vegetative stage

JOURNAL Unpublished (2002)  
 COMMENT Contact: Tadasu Shin-i  
 Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshini@genes.nig.ac.jp.  
 Location/Qualifiers

FEATURES  
 source 1..584  
 /organism="Dictyostelium discoideum"  
 /strain="AX4"  
 /db\_xref="taxon:44689"  
 /clone\_lib="ddv49k01"  
 /clone\_lib="Dictyostelium discoideum cDNA library, VF"  
 /sex="mat A"  
 /dev\_stage="Growth phase"  
 BASE COUNT 214 a 60 c 89 g 221 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.61 Length: 584  
 Score: 91.00 Matches: 50  
 Percent Similarity: 37.39% Conservative: 33  
 Best Local Similarity: 22.52% Mismatches: 79  
 Query Match: 6.08% Indels: 60  
 DB: 13 Gaps: 11

US-10-062-624-40 (1-293) x BJ423418 (1-584)  
 Qy 9 IleIleAsnThrValLeuValCysLeu-----LeuSerLeuProAsnIleSer 24  
 Db 31 GTCAATCAATATGTTGTGTATATGTAATAATATGTTATATATATATATATATATCT 90  
 Qy 25 SerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyrIleSerGlyGln 44  
 Db 91 ATTTTAAACAGATCAACAATAAGAAAA-----ATGAAAGATCA 132  
 Qy 45 TyrLysProSerValPheSerAsnPheSerValLysGluThrAsnValIleThr 64  
 Db 133 GAAAAATCAATCATCTGTTGTTAGTAATAACAACAACAAGATGTAATATCATCAGT 192  
 Qy 65 LysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLysThrAspAlaSer 84  
 Db 193 TCAAAATGAAGTGTGTTAAAGAA-----GAAAAATAAA----- 225  
 Qy 85 ValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaValPheGlnAspAsn 104  
 Db 226 -----GGACATCAAGAAATCTTATTAAGTTTAAATCTA----- 261  
 Qy 105 SerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArg----- 121  
 Db 262 -----TCAGTCATTTATGTTTATGATTTTCAACTCGTTTATTCTCA 303  
 Qy 122 ---ValGluIleGluGlySerTyrGluGluPheAsp----- 132  
 Db 304 GTATTAGCTTATGAAGTGTATTATTCATGAATTTGATCCATATTTTAATTATAGATCAACA 363  
 Qy 133 -----ValLysAsnProGlyTyrThrLeu-----SerAspAlaTyrArgTyr 147  
 Db 364 ATATATCTTGTTCAGAGGTTTTTATAATTTTAAATGCTTTGATGAAGAGCATGG 423  
 Qy 148 PheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGluLysValSerAsn 167  
 Db 424 TATCCATTAGCAGCTATTGTAGTGGTACAACTTAC---CCAGGTTTAAATGGCAACGCA 480  
 Qy 168 SerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSerValIleValAsn 187  
 Db 481 AGTTTAGTTTCATGCTCATTC-----AATTCATGAATATTACAGTTAATATTAGAAAT 534  
 Qy 188 ValCysTyrAspPheSerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAla 207  
 Db 535 GTATGT-----GTATGTTATCACCATGGTTTGCATCAATAACA 573  
 Qy 208 GlyVal 209  
 Db 574 GCAATG 579  
 RESULT 29  
 BJ332457 590 bp mRNA linear EST 05-MAR-2002  
 LOCUS BJ332457 Dictyostelium discoideum cDNA library, AF Dictyostelium  
 DEFINITION discoideum cDNA clone dda39n05 5', mRNA sequence.

ACCESSION BJ332457  
 VERSION BJ332457.1 GI:19162587  
 KEYWORDS  
 SOURCE EST.

ORGANISM Dictyostelium discoideum.  
 REFERENCE 1 (bases 1 to 590)  
 AUTHORS Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.  
 TITLES Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.  
 Full length cDNA of Dictyostelium discoideum at the aggregation stage

JOURNAL Unpublished (2002)  
 COMMENT Contact: Tadasu Shin-i  
 Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshini@genes.nig.ac.jp.



```

QY 144 AlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGlu 163
DB 405 GAAAGAGCATGTTATCCATTAGGACGT---ATTGTAGGTGGTACCAATTTACCAGGTTTA 461
QY 164 LysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSer 183
DB 462 ATGGCAACAGCAAGTTAGTTCATTGGTCATTG-----AATTCATTGAATATTACAGTT 515
QY 184 ValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSerProTyrIle 203
DB 516 AATATTAGAATGTAATG-----GTATTGTTATCACCACCATGGTTT 554
QY 204 CysGlyGlyAlaGlyValAspAlaIleGluPhe 214
DB 555 GCATCAAAATACAGCAATGGTAACTTATAAAATTT 587

RESULT 31
LOCUS BJ422368
DEFINITION BJ422368 Dictyostelium discoideum cDNA library, VF Dictyostelium
ACCESSION BJ422368
VERSION BJ422368
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE 1 (bases 1 to 591)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the vegetative
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
source
1. .591
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone_lib="ddv45014"
/sex="mat A"
/dev_stage="Growth phase"
BASE COUNT 216 a 62 c 90 g 223 t
ORIGIN

Alignment Scores:
Pred. No.: 0.623 Length: 591
Score: 91.00 Matches: 53
Percent Similarity: 36.80% Conservative: 32
Best Local Similarity: 22.94% Mismatches: 86
Query Match: 6.08% Indels: 60
DB: 13 Gaps: 11

US-10-062-624-40 (1-293) x BJ422368 (1-591)

QY 5 LeuLysPheThrIleIleAsnThrValLeuValCysLeu-----LeuSerLeu 20
DB 11 TTGTTATTTTAGTCAATCAATATTTGTGTATATATATAATATGTTATATTAGTGA 70
QY 21 ProAsnIleSerSerLysValAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40
DB 71 TAAATATATATCTATTTTAAACAGATCAACATTAAGAAAAA----- 112
QY 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60
DB 113 ATGMAAGATCAGAAAAATCAAGTACATCTGTTGTAGTAAATAACAACAACAGAGTGA 172

```

```

QY 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
DB 173 AATATCATCATGTTCAATGAAGTTGGTGTAAAGAA-----GAAATAAA 217
QY 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaVal 100
DB 218 -----GGACATCAAGAAATCTTATTAAAAAGTTTTTAATCTA 253
QY 101 PheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThr 120
DB 254 -----TCAGTCATTATGTTTGTAGCATTTTCACT 283
QY 121 Arg-----ValGluIleGluGlySerTyrGluGluPheAsp----- 132
DB 284 CGTTTATTCTCAGTATTACGTTATGAAGTGTTATTCATGAATTTGATCCATATTTTAAT 343
QY 133 -----ValLysAsnProGlyTyrThrLeu-----SerAsp 143
DB 344 TATAGATCAACAATATATCTGTTCAAGAGTTTTTATAATTTTTTAAATGGTTGAT 403
QY 144 AlaTyrArgTyrPheAlaLeuAlaLeuGluMetLysGlyAsnSerPheThrProLysGlu 163
DB 404 GAAAGAGCATGTTATCCATTAGGACGT---ATTGTAGGTGGTACAAATTTACCAGGTTTA 460
QY 164 LysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSer 183
DB 461 ATGGCAACAGCAAGTTAGTTCATTGGTCATTG-----AATTCATTGAATATTACAGTT 514
QY 184 ValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSerProTyrIle 203
DB 515 AATATTAGAATGTAATG-----GTATTGTTATCACCACCATGGTTT 553
QY 204 CysGlyGlyAlaGlyValAspAlaIleGluPhe 214
DB 554 GCATCAAAATACAGCAATGGTAACTTATAAAATTT 586

RESULT 32
LOCUS BJ425564
DEFINITION BJ425564 Dictyostelium discoideum cDNA library, VF Dictyostelium
ACCESSION BJ425564
VERSION BJ425564.1 GI:19342271
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE 1 (bases 1 to 591)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the vegetative
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
source
1.591
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone_lib="ddv56k03"
/sex="mat A"
/dev_stage="Growth phase"
BASE COUNT 215 a 60 c 91 g 225 t
ORIGIN

Alignment Scores:

```



Pred. No.:	0.623	Length:	591
Score:	91.00	Matches:	53
Percent Similarity:	36.80%	Conservative:	32
Best Local Similarity:	22.94%	Mismatches:	85
Query Match:	6.08%	Indels:	60
DB:	13	Gaps:	11

US-10-062-624-40 (1-293) x BJ425564 (1-591)

QY	5	LeuLysPheThrIleAlaSerValLeuValCysLeu-----LeuSerLeu 20
Db	15	TTGTTATTAGTCAATCAATATTGTGTGTATATGTAATAAATGGTATATTAGTGTA 74
QY	21	ProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyThrGlyLeuTyR 40
Db	75	TAATATATATCTATTTTAAACAGATCAACAAATAAGAAAAA----- 116
QY	41	IleSerGlyGlnTyLysProSerValSerValPheSerAsnPheSerValLysGluThr 60
Db	117	ATGAAAGATCAGAAAATCAAGTACATCTGTTGTTAGTAATAACAACAACAGATGTA 176
QY	61	AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
Db	177	AATATCATCAGTTCAAAATGAAGTTGGTGTAAAGAA-----GAAATATA 221
QY	81	ThrAspAlaSerValGlyLysSerAsnProSerAsnPheThrLeuProTyThrAlaVal 100
Db	222	-----GGACATCAAGAAATCTTATTAAAGATTTTAAATCTTA 257
QY	101	PheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyThrPheAlaGluGlyThr 120
Db	258	-----TCAGTCATTTATGTTTGTAGCATTTTCAACT 287
QY	121	Arg-----ValGluIleGluLysSerTyrgluGluPheAsp----- 132
Db	288	CGTTTATCTCAGTATTACGTTATGAAGTGTTATTATCATGAATTTGATTCATATTTAAT 347
QY	133	-----ValLysAsnProGlyTyThrLeu-----SerAsp 143
Db	348	TATAGATCAACAATATATCTGTTCAAGAGGTTTTTATAATTTTAAATTCGTTTGTAT 407
QY	144	AlaTyArgTyThrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGlu 163
Db	408	GAAAGACGATGGTATCCATTAGGACGT---ATTGAGTGGTATCAATTTACCCAGGTTTA 464
QY	164	LysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSer 183
Db	465	ATGGCAACAGCAAGTTTAGTTCATTTGGTCATG-----AATTCATTGAATATTACAGTT 518
QY	184	ValIleValAsnValCysTyRAspPheSerLeuAsnLeuSerIleSerProTyIle 203
Db	519	AATATTAGAAATGTATGT-----GTATGTTATCACCATGTTT 557
QY	204	CysGlyGlyAlaGlyValAspAlaIleGluPhe 214
Db	558	GCATCAAAATACAGCAATGCTACCTATAAAATTT 590

<b>RESULT 33</b>	BJ425783	LOCUS	BJ425783	591 bp	mRNA	linear	EST 11-MAR-2002
<b>DEFINITION</b>			BJ425783 Dictyostelium discoideum CDNA library, VF				Dictyostelium discoideum CDNA clone dv56g23 5', mRNA sequence.

COMMENT

Contact: Tadasu Shin-i  
 Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshini@genes.nig.ac.jp.

Alignment Scores:		
Pred. No.:	0.623	Length:
Score:	91.00	Matches:
Percent Similarity:	36.80%	Conservative:
Best Local Similarity:	22.94%	Mismatches:
Query Match:	6.08%	Indels:
DB:	13	Gaps:
		591
		53
		32
		86
		60
		11

US-10-062-624-40 (1-293) x BJ425783 (1-591)

Qy	5	LeuLysPheThrIleIleAenThrValLeuValCysLeu-----LeuSerLeu	20
Db	8	TTGTGTAATTTAGTCAATCAATATTGTGTGTATATGAATAAAATGGTGATATAGTAGTA	67
Qy	21	ProAsnIleSerSerLysAlaIleAsnAsnAenAlaLysLysTy-TyrGlyLeuTyr	40
Db	68	TAAATATATATCTATTTTTAAACAGATCACAAATAGAATAAAA-----	109
Qy	41	IleSerGlyGlnTyrLysProSerValSerValPheSerAenPheSerValLysGluThr	60
Db	110	ATGAAAGATCAGAAAAATCAAGTACATCTGTTGTAGTAATAACAAAACAACAGATGTA	169
Qy	61	AsnValIleThrLysAenLeuIleAlaLeuLysLysAspValaspSerIleGluThrLys	80
Db	170	AATCATCATCGTCAAATGAAGTGGGTGTTAAAGAA-----GAAATAAA	214
Qy	81	ThrAspAlaSerValGlyIleSerAsnProSerAenPheThrIleProTyrThrAlaVal	100
Db	215	-----GGACATCAACAAGATCTTTATTAAAGATTTTAATTCCTA	250
Qy	101	PheClnAspAsnSerValAenPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThr	120
Db	251	-----TCAGTCATTTATGTTTATAGCATTTTCCAAT	280
Qy	121	Arg-----ValGluIleGluGlySerTyrGluGluPheAsp-----	132
Db	281	CGTTTATTCAGTATTACGTTATGAAGAGTGTATTCAATGATTGATCCATATTTAAT	340
Qy	133	-----ValLysAsnProGlyGlyTyrThrLeu-----SerAsp	143
Db	341	TATAGATCAACAATATATCTGTTTCAAGAAGGTTTTTATAATTTTTTAAATGGTGTGAT	400
Qy	144	AlaTyrArgTyrPheAlaLeuAlaArgLumetLysGlyAenSerPheThrProLysGlu	163
Db	401	GAAAGACATGGGTATCCATTAGGACGT--ATTGTAGTGTGTACAAATTTACCACGGTTTA	457
Qy	164	LysValSerAenSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSer	183
Db	458	ATGCACAACACCAAGTTTAGTTCATTGTCATTG-----AATTCATTCAATATTACAGTT	511
Qy	184	ValIleValAenValCysTyrAspPheSerLeuAenAenLeuSerIleSerProTyrIle	203
Db	512	AATATTAGAANTGTATGT-----GTATGTTTATCCACATGGTGGTT	550





```

ORGANISM Dictyostelium discoideum
REFERENCE Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
AUTHORS 1 (bases 1 to 597)
TITLE Full length cDNA of Dictyostelium discoideum at the aggregation
        stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
        National Institute of Genetics
        1111 Yata, Mishima, Shizuoka 411-8540, Japan
        Tel: 81-559-81-6856
        Fax: 81-559-81-6855
        Email: tsinin@genes.nig.ac.jp.
FEATURES
        source
        1..597
        /organism="Dictyostelium discoideum"
        /strain="AX4"
        /db_xref="taxon:44689"
        /clone="dda41115"
        /clone_lib="Dictyostelium discoideum cDNA library, AF"
        /sex="mat A"
        /dev_stage="Aggregation stage"
BASE COUNT 219 a _62 c 91 g 225 t
ORIGIN
Alignment Scores:
Pred. No.: 0.634 Length: 597
Score: 91.00 Matches: 53
Percent Similarity: 36.80% Conservative: 32
Best Local Similarity: 22.94% Mismatches: 86
Query Match: 6.08% Indels: 60
DB: 13 Gaps: 11
US-10-062-624-40 (1-293) x BJ333160 (1-597)
QY 5 LeuLysPheThrIleAsnThrValLeuValCysLeu-----LeuSerLeu 20
DB 15 TTGTTATTCTCAATCAATATTGTTGTATATTAATAATTGTTATATTAGTGTA 74
QY 21 ProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40
DB 75 TAATATATATCTATTTTAAACAGATCAACATTAAGAAAAA-----116
QY 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60
DB 117 ATGAAAGATCAGAAAAATCAAGTACATCTGTTAGTAAACAAACAACAGATGTA 176
QY 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
DB 177 AATATCATCAGTTCAATGAAGTGGTGTAAAGAA-----GAAATATAA 221
QY 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaVal 100
DB 222 -----GGACATCAAGATCTTATTAAGATTTTAAATCTA 257
QY 101 PheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyThrPheAlaGluGlyThr 120
DB 258 -----TCAGTCATTTATGTTTATAGCATTTTCACT 287
QY 121 Arg-----ValGluIleGluGlySerTyrGluGluPheAsp-----132
DB 288 CGTTTATCTCAGTATTGCTTGAAGTGTATTTCATGAATTTGATCCATATTTTAAAT 347
QY 133 -----ValLysAsnProGlyTyrThrLeu-----SerAsp 143
DB 348 TATAGATCAACAATATCTCTGTTCAAGAGGTTTATTAATTTTAAATGGTTGAT 407
QY 144 AlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGlu 163
DB 408 GAAAGAGCATGTTATCCATTAGGACGT---ATTGTAGTGGTGACACATTTACCACGTTTA 464
QY 164 LysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSer 183

```

```

DB 465 ATGCAACAGCAAGTTAGTTTCATGTCATG-----AATTCATTGAATATTACAGTT 518
QY 184 ValIleValAsnValCysTyrAspPheSerLeuAsnLeuSerIleSerProTyrIle 203
DB 519 AATATTAGAAATGTATGT-----GTATTGTTATCACCATGGTTT 557
QY 204 CysGlyGlyAlaGlyValAspAlaIleGluPhe 214
DB 558 GCATCAAAATACAGCAATGGTAACTAATAAATTT 590
RESULT 38
LOCUS BJ335009
DEFINITION BJ335009 Dictyostelium discoideum cDNA library, AF Dictyostelium
        discoideum cDNA clone dda48h17 5', mRNA sequence.
ACCESSION BJ335009
VERSION BJ335009.1 GI:19165139
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 597)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the aggregation
        stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
        Center For Genetic Resource Information
        National Institute of Genetics
        1111 Yata, Mishima, Shizuoka 411-8540, Japan
        Tel: 81-559-81-6856
        Fax: 81-559-81-6855
        Email: tsinin@genes.nig.ac.jp.
FEATURES
        source
        1..597
        /organism="Dictyostelium discoideum"
        /strain="AX4"
        /db_xref="taxon:44689"
        /clone="dda48h17"
        /clone_lib="Dictyostelium discoideum cDNA library, AF"
        /sex="mat A"
        /dev_stage="Aggregation stage"
BASE COUNT 219 a _62 c 91 g 225 t
ORIGIN
Alignment Scores:
Pred. No.: 0.634 Length: 597
Score: 91.00 Matches: 53
Percent Similarity: 36.80% Conservative: 32
Best Local Similarity: 22.94% Mismatches: 86
Query Match: 6.08% Indels: 60
DB: 13 Gaps: 11
US-10-062-624-40 (1-293) x BJ335009 (1-597)
QY 5 LeuLysPheThrIleIleAsnThrValLeuValCysLeu-----LeuSerLeu 20
DB 15 TTGTTATTCTCAATCAATATTGTTGTATATTAATAATTGTTATATTAGTGTA 74
QY 21 ProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40
DB 75 TAATATATATCTATTTTAAACAGATCAACATTAAGAAAAA-----116
QY 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60
DB 117 ATGAAAGATCAGAAAAATCAAGTACATCTGTTAGTAAACAAACAACAGATGTA 176
QY 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
DB 177 AATATCATCAGTTCAATGAAGTGGTGTAAAGAA-----GAAATATAA 221
QY 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaVal 100

```



Email: tshini@genes.nig.ac.jp.

FEATURES  
source

Location/Qualifiers  
1..597  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone="ddv45d10"  
/clone\_lib="Dictyostelium discoideum cDNA library, VF"  
/sex="mat A"  
/dev\_stage="Growth phase"

BASE COUNT 219 a \_62 c 91 g 225 t  
ORIGIN

## Alignment Scores:

Pred. No.:	0.634	Length:	597
Score:	91.00	Matches:	53
Percent Similarity:	36.80%	Conservative:	32
Best Local Similarity:	22.94%	Mismatches:	86
Query Match:	6.08%	Indels:	60
DB:	13	Gaps:	11

US-10-062-624-40 (1-293) x BJ422231 (1-597)

```
QY 5 LeuLysPheThrIleLeuValCysLeu-----LeuSerLeu 20
Db 15 TTGTTATTTAGTCAATCAATTTGTTGTTATTAATAATTTGTTATTTAGTGA 74
QY 21 ProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyrGlyLeuTyr 40
Db 75 TAAATATATCTATTTTAAACAGATCAACATAAGAAAAA-----116
QY 41 IleSerGlyLysProSerValSerValPheSerAsnPheSerValLysGluThr 60
Db 117 ATGAAAGATCAGAAAAATCAAGTACATCTGTTGTTAGTAATAACAAACAAGATGA 176
QY 61 AsnValIleThrLysLeuLeuAlaLeuLysLysAspValAspSerIleGluThrLys 80
Db 177 AATATCATCAGTTCAATGAAAGTTGGTGTAAAGAA-----GAAATAAA 221
QY 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaVal 100
Db 222 -----GGACATCAAGAAATCTTATTAAAGTTTTTAATTTCTA 257
QY 101 PheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThr 120
Db 258 -----TCAGTCATTATGTTTACGATTTTCAACT 287
QY 121 Arg-----ValGluIleGluGlySerTyrGluGluPheAsp-----132
Db 288 CGTTTATCTCAGTATAGTTATGAAAGTGTATTTCATGAATTTGATCCATATTTAAT 347
QY 133 -----ValLysAsnProGlyTyrThrLeu-----SerAsp 143
Db 348 TATAGATCAACAATATCTCTGTTCAAGAGGTTTTTATAATTTTAAATTTGGTTGAT 407
QY 144 AlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGlu 163
Db 408 GAAAGAGCATGTTATCCATTAGGACGT---ATTGTAGTGGTGACAAATTTACCCAGGTTTA 464
QY 164 LysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSer 183
Db 465 ATGGCAACAGCAAGTTTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 518
QY 184 ValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSerProTyrIle 203
Db 519 AATATTAGAAATGTTATGTT-----GTATTGTTATCACCATGGTTT 557
QY 204 CysGlyGlyAlaGlyValAspAlaIleGluPhe 214
Db 558 GCATCAATACAGCAATGTTACCTATTAATTT 590
```

Search completed: July 3, 2003, 23:30:34

Job time : 1104 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - nucleic search, using frame\_plus\_p2n model

Run on: July 3, 2003, 23:12:53 ; Search time 46 Seconds  
(without alignments)  
1953.398 Million cell updates/sec

Title: US-10-062-624-40  
Perfect score: 1496  
Sequence: 1 MNMKKFTIINTVLVCLSL.....AVATLNIYFGEGIGARLTF 293

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 segs, 1533381 residues  
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO/US10062624/runat\_30062003\_091230\_243332/app\_query.fasta\_1.455  
-DB=Issued\_Patents\_NA -QWPT=fastcap -SUFFIX=rn1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTPMT=pico -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10062624.qcgn.1.1.32.@runat\_30062003\_091230\_24332 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGESQUERY -NEG\_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1496	100.0	879	4	US-09-660-587-39
2	490.5	32.8	840	4	US-09-660-587-41
3	468.5	31.3	843	4	US-08-953-326-10
4	459	30.7	864	3	US-08-733-230-1
5	446.5	29.8	864	4	US-08-953-326-1
6	446.5	29.8	840	4	US-09-660-587-5
7	446.5	29.8	840	4	US-09-261-358A-5
8	443.5	29.6	1607	4	US-09-660-587-1
9	443.5	29.6	1607	4	US-09-261-358A-1
10	443.5	29.6	1607	4	US-09-648-520E-47
11	443.5	29.6	1607	4	US-09-201-458-1
12	441.5	29.5	837	4	US-08-953-326-9

13	441	29.5	849	4	US-09-660-587-3	Sequence 3, Appli
14	441	29.5	849	4	US-09-261-358A-3	Sequence 3, Appli
15	437	29.2	819	4	US-09-660-587-45	Sequence 45, Appli
16	427	28.5	861	4	US-08-953-326-8	Sequence 8, Appli
17	426	28.5	864	4	US-08-953-326-12	Sequence 12, Appli
18	423.5	28.3	828	4	US-09-660-587-43	Sequence 43, Appli
19	422	28.2	830	4	US-08-953-326-11	Sequence 11, Appli
20	414.5	27.7	849	4	US-09-648-520E-48	Sequence 48, Appli
21	409.5	27.4	842	3	US-08-733-230-3	Sequence 3, Appli
22	409.5	27.4	842	3	US-08-953-326-3	Sequence 3, Appli
23	294	19.7	849	3	US-08-733-230-5	Sequence 5, Appli
24	294	19.7	849	4	US-08-953-326-5	Sequence 5, Appli
25	239.5	16.0	1278	4	US-09-288-339-3	Sequence 3, Appli
26	236	15.8	1840	4	US-09-066-047-26	Sequence 26, Appli
27	236	15.8	1843	4	US-09-066-047-12	Sequence 12, Appli
28	217.5	14.5	3435	4	US-09-066-046-28	Sequence 28, Appli
29	217.5	14.5	3435	4	US-09-066-047-14	Sequence 14, Appli
30	197.5	13.2	1242	4	US-09-288-339-1	Sequence 1, Appli
31	196.5	13.1	1176	4	US-09-288-339-5	Sequence 5, Appli
32	196	13.1	1345	4	US-08-975-762-1	Sequence 1, Appli
33	196	13.1	1345	4	US-08-821-324-1	Sequence 1, Appli
34	196	13.1	1345	4	US-08-255-028-1	Sequence 1, Appli
35	196	13.1	1345	4	US-09-106-582-1	Sequence 1, Appli
36	196	13.1	1980	4	US-09-235-028-84	Sequence 84, Appli
37	187	12.5	1919	4	US-08-975-762-40	Sequence 40, Appli
38	187	12.5	1919	4	US-09-235-028-40	Sequence 40, Appli
39	187	12.5	1919	4	US-09-106-582-40	Sequence 40, Appli
40	173	11.6	399	4	US-08-953-326-13	Sequence 13, Appli
41	164.5	11.0	3998	4	US-09-066-046-5	Sequence 5, Appli
42	144.5	9.7	3435	4	US-09-066-046-28	Sequence 28, Appli
43	144.5	9.7	3435	4	US-09-066-047-14	Sequence 14, Appli
44	133.5	8.9	1132	4	US-08-975-762-2	Sequence 2, Appli
45	133.5	8.9	1132	4	US-08-821-324-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-09-660-587-39  
Sequence 39, Application US/09660587  
Patent No. 6392023  
GENERAL INFORMATION:  
APPLICANT: Walker, David H.  
APPLICANT: McBridge, Jere W.  
APPLICANT: Yu, Xue-Jie  
TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
FILE REFERENCE: D6152CIP2  
CURRENT APPLICATION NUMBER: US/09/660,587  
CURRENT FILING DATE: 2000-09-12  
PRIOR APPLICATION NUMBER: 09/261,358  
PRIOR FILING DATE: 1999-03-03  
NUMBER OF SEQ ID NOS: 46  
SEQ ID NO 39  
LENGTH: 879  
TYPE: DNA  
ORGANISM: Ehrlichia canis  
FEATURE:  
OTHER INFORMATION: nucleic acid sequence of E. canis p28-1  
US-09-660-587-39

Alignment Scores:  
Pred. No.: 1.2e-175  
Score: 1496.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 100.00%  
DB: 4  
Gaps: 0  
US-10-062-624-40 (1-293) x US-09-660-587-39 (1-879)

Qy 1 MetasnsnylsleuysphethrileileksnrthValleuValCysLeuSerIeu 20  
|||||

```

Db      1 ATGAATAATAAAGTCAATTTACTATAATAAACACAGTATTAGTATGCTATTGTCAATTA 60
Qy      21 ProaenlleSerSerSerlyValalleaenaaenallalylygTyTyTyGlyLeuTy 40
Db      61 CCTAATAATAATCTCTCAAGGCGCAATAAACATTAACGCTAAAGTACTACGATTAAT 120
Qy      41 lleserGlyGlyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 60
Db      121 ATCAGTGCACATATTAACCCAGTGTCTGTTTCAGTAATTTTTCAGTTAAAGAAC 180
Qy      61 AenValilleThlysaenleullealeuLylysaenVallePheSerlleGlyLeuTy 80
Db      181 AATGCATTAATAAAGCTTATAGCTTAAAGAAAGTGTGACTTATTAAGAACCAAG 240
Qy      81 ThrAaPalaservalGlylleSerleuPheSerleuPheSerleuPheSerleuPheSer 100
Db      241 ACTGATGCCAGTGAAGTATTAAGTAAACCATTAATTTACTATATCCCTATACACTGA 300
Qy      101 PheGlnAaPaSerSerValaenPheAenGlyThrlleGlyTyTyThPheAagluGlyTh 120
Db      301 TTTCAGATTAATCTCTCAATTTCAATGGAATTTGTTTACCTTTCGTAAGTACA 360
Qy      121 ArgValagluGlyGlyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 140
Db      361 AAGATTGAATGAAGGTTCTTATGAGAAATTTGAATTTAAACCTGAGGCTATACA 420
Qy      141 LeuSerAaPaLleTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 160
Db      421 CTAAAGTGAAGCTATAGCTATTTTGCATTAAGACGGAATTAAGATTAAGTTTACA 480
Qy      161 ProLyGluLyValSerAenSerllePheThsthrValleAaArgAaPaLleLeuSer 180
Db      481 CCTAAAGAAAGTTCTCTAATAGTATTTTCAACGCTAATGAAGATTAAGTATCT 540
Qy      181 lleleSerValilleValaenValCyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 200
Db      541 ATTAATATCTGTATAGTAAATGTTGCTACGATTTCTCTTGAACAATTTTCAATATCG 600
Qy      201 ProTyThrlleCyGlyGlyValaValaPaalleGluPhePaPaValleuThlle 220
Db      601 CCTTACATTAATGAGGAGACAGGAGTAAAGTAAATTTCTTGCATTAATTAACAT 660
Qy      221 LysPheAaLleTyGlnSerlySerlyGlylleAaTyTyTyTyTyTyTyTyTyTyTy 240
Db      661 AAGTTTGCATTAACAAAGCAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 720
Qy      241 PheAaSerLeuTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 260
Db      721 TTTGCTAGTTTATTAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 780
Qy      261 HisValaIaGluLeuAaSerllePheProlyslleThSerAaValaAaThreAenlle 280
Db      781 CAGTGTGCAAGTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 840
Qy      281 GlyTyTyPheGlyGlyGlylleGlyAaArgLeuThPhe 293
Db      841 GGTATTTTGAAGTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 879

```

```

; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 41
; LENGTH: 840
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: nucleic acid sequence of E. canis p28-2
; US-09-660-587-41

Alignment Scores:
Pred. No.: 2,44e-51 Length: 840
Score: 490.50 Matches: 114
Percent Similarity: 56.35% Conservative: 59
Best Local Similarity: 37.13% Mismatches: 85
Query Match: 32.79% Indels: 49
DB: 4 Gaps: 12

US-10-062-624-40 (1-293) x US-09-660-587-41 (1-840)
Qy      6 LysPheThrllelleAenThrValleuValCysLeuLeuSer---LeuProAaenlleSer 24
Db      10 AAGAAATTCATAGTAAGAGGCGCTTATCTCAATTAATGCTAATCTTACATACAGCT 69
Qy      25 SerSerlyAaLle-----AaAaAaAaAaLylyTyTyTyTyTyTyTyTyTyTyTy 40
Db      70 TTTCAGATCTCTGAGTTCAGAACTAATGATTAACAAAGAA-----GGCTTCTAC 120
Qy      41 lleserGlyGlyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 59
Db      121 ATTAATGCAAGTCAATCCAAAGTAAATCAACACTTGAAGAAATTTCTGCTGAAGAACT 180
Qy      60 -----ThraenValilleThlysaenleullealeuLylysaenValleAaPa 75
Db      181 CCTAATAATGAACAAATTTCTCACTAAAGAAATTTTTCGACCTTAAGAAAGATGCTAT 240
Qy      76 SerlleGluThrllysthrAaPaLleSerValGlylleSerAaenPheSerleuThrlle 95
Db      241 -----ATAACAATAAAGACGATTTTATCAAGA 267
Qy      96 ProTyThrlleAaVal---PheGlnAaPaSerValaen---PheAenGlyThrlleGly 113
Db      268 GTAGCTCCAGGCAATGATTTTCAAAATTAATTAATCAAGATTTTCAAGAAATTTGCT 327
Qy      114 TyThrPheAaGluGlyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 133
Db      328 TACTCTAAG---GACGCAACCAAGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 384
Qy      134 LysAaenPro-----GlyGlyTyThrlleuSerAaPaLleTyTyTy 146
Db      385 AAAAAACCGATTAACATGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 429
Qy      147 TyThrPheAaLleuAaAaGluMetLeGlyAaenPheThrlleProlyGluLyValSer 166
Db      430 CATTTCATTAATTTCTG-----AAAGATGAATGGA 462
Qy      167 AaenSerllePheThsthrValleAaArgAaPaLleLeuSerllelleSerValilleVal 186
Db      463 GATCAGCAATAT---GTAAGTCTTAATAATTAACGCAATATTTATGCTATTAAGTGT 519
Qy      187 AaenValCyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 206
Db      520 AATACCTGCTAATGACATTAACAGTGAAGAGTAACTTTCCGACATTAATGATGTCAGGT 579
Qy      207 AlaGlyValaAaPaalleGluPhePaPaValleuThlleThlysaenPheAaLleGlnSer 226
Db      580 ATAGAGAGAGATCTTATCACTATTAATTAAGACCTCAATTAATTAATTAATTAATTA 639
Qy      227 LysLeuGlylleAaTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 246
Db      640 AAAAAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 699
Qy      247 HisLyValMetGlyAaenGlnPheLyAaenleuAaenValGlnHisValaIaGluLeuAa 266

```



Db 700 CATGCCGTTATGGTAATTAATTGAGAAAGTACTGTAACTCCTGTAGTATTAAT 759  
Qy 267 SerleuprolylethrSerAlaValAlaThrLeuAenIleGlyTyrPheGlyGly 286  
Db 760 GATGCTCTCCAAACCACTCTGCTTCACTTGAAGTGTGACTTTGGCGAGAA 819  
Qy 287 IleGlyAlaArgLeuThrPhe 293  
Db 820 ATGGGAATGAGGTTCACTTC 840

RESULT 3  
US-08-953-326-10  
Sequence 10, Application US/08953326  
Patent No. 6251872  
GENERAL INFORMATION:  
APPLICANT: Barbec, Anthony F.  
APPLICANT: Ganca, Roman R.  
APPLICANT: McGuire, Travis C.  
APPLICANT: Burridge, Michael J.  
APPLICANT: Nyika, Aceme  
APPLICANT: Rurangirwa, Fred R.  
TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of  
TITLE OF INVENTION: Animals and Humans  
FILE REFERENCE: UF-167C1  
CURRENT APPLICATION NUMBER: US/08/953,326  
EARLIER FILING DATE: 1997-10-17  
EARLIER APPLICATION NUMBER: 08/953,326  
EARLIER FILING DATE: 1997-10-17  
EARLIER APPLICATION NUMBER: 08/733,230  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 10  
LENGTH: 843  
TYPE: DNA  
ORGANISM: Ehrlichia chaffeensis  
US-08-953-326-10

Alignment Scores:  
Pred. No.: 1,296-48 Length: 843  
Score: 468.50 Matches: 115  
Percent Similarity: 56.03% Conservative: 57  
Best Local Similarity: 37.46% Mismatches: 94  
Query Match: 31.32% Indels: 41  
Gaps: 14

US-10-062-624-40 (1-293) x US-08-953-326-10 (1-843)

Qy 1 MetAsnAenLYLeuLYPheThrIleIleAenThrValLeuValCysLeuLeuSer--- 19  
Db 3 ATGAAATTCGAA--AAATTT---TTTATACAACTACATTAATGATCCCTAATGCTTC 56  
Qy 20 LeuProAenIleSerSerSerLYAlaIleAenAenAlaLYLeuTyrTyrGlyLeu 39  
Db 57 TTACTGGAATATATCTTTCTTGATGCAGTACGAAAGC---AANGTGGGATTAATTC 113  
Qy 40 TyrIleSerGlyGlyTyrLYPheProSerValSerValPheSerAenPheSerValIleGly 59  
Db 114 TATATCACTGGGAATATGATGACCAAGTCTTTCACATTTGCGATTTCTCTGCTTAACAG 173  
Qy 60 ThrAsnValIleThrLYAsnLeuIleAlaLeuLYAspValAspSerIleGluThr 79  
Db 174 GAAAGAAATACAACTACGAGTATTTGATTAAGCAAGATTGGAGTGC----- 224  
Qy 80 LysThrAspAlaSerValGlyIleSerAenProSerAen---PheThrIlePro---Tyr 97  
Db 225 -----AGCACATATCTTAATAAATCTCCAGAAAATACATTTAAGTTCACAAATAT 275  
Qy 98 ThrAlaValIleGlnAspAenSer---ValAsnIleAenGlyThrIleGlyTyrThrPhe 116  
Db 276 TCATTAAATATGAAATAATCACTTCTAGTGTTCAGAGAGCTGTGGTTAT---TTA 332

Qy 117 AlaGluGlyThrArgValGluIleGlyIleSerTyrGluGluPheAspValLYAsnPro 136  
Db 333 ATGAATGCTCCAAAGAAATAGTTGAATGATGCCATATGAACATTTGATGAAAAACAG 392  
Qy 137 GlyGlyTyrThrLeuSerAenPalaTyrArgTyrPheAlaIleValAlaArgIleMetGly 156  
Db 393 GGTAAATACATTAAGAACGATGCTCACAAATATATATGCTTTAACCCATAACAGTGGGGA 452  
Qy 157 AsnSerPheThrProLYSGlyValSerAenSer-----IlePheIleThr 172  
Db 453 -----AACCTAAGCAATGACAGTGAATAGTTGTTT----- 485  
Qy 173 ValMetArgAenAenPGLYLeuSerIleIleSerValIleValAenValCysTyrAspPhe 192  
Db 486 -----CTAATAAATGAGAGCTCTTGATATATCACTTATGTTGAATGATGCTATGATGTA 542  
Qy 193 SerLeuAenAenLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIle 212  
Db 543 ATAAGTGAAGGAATACCTTCTCTCTTACATATGTCAGAGTGTGGTATGATTAATA 602  
Qy 213 GluPhePheAspValIleuHleIleLYPheAlaTyrGlnSerLYLeuGlyIleAlaTyr 232  
Db 603 TCCATGTTTGAAGCTATTAACCTTAATAATTTCTTATACAGAAAGTTAAGTTATAC 662  
Qy 233 SerLeuProSerAenIleSerLeuPheAlaSerLeuTyrTyrHISLYValMetGlyAsn 252  
Db 663 TCCATTAAGCCAGAGAGCTCTGTTTCTTGTTGTCGACATTTTCTTAAGTATAGGGAAT 722  
Qy 253 GlnPheLYAsnLeuAenValGlnHleValAlaGluLeuAlaSerIleProLYIleThr 272  
Db 723 GAATTCAAGAGATATTCCT-----GCTATGATATACCAAGTACCTCA 761  
Qy 273 Ser-----AlaValAlaThrLeuAenIleGlyTyrPheGlyGly 286  
Db 762 ACTCTCAGAGTAATCACTTACTATATGATACACTAAGTATGACCACTTTGAGTGA 821  
Qy 287 IleGlyAlaArgLeuThrPhe 293  
Db 822 CTGGAGAGAGTTTAACTTT 842

RESULT 4  
US-08-733-230-1  
Sequence 1, Application US/08733230  
Patent No. 6025338  
GENERAL INFORMATION:  
APPLICANT: Barbec, Anthony F.  
APPLICANT: Ganca, Roman Reddy  
APPLICANT: McGuire, Travis C.  
APPLICANT: Burridge, Michael J.  
APPLICANT: Nyika, Aceme  
APPLICANT: Rurangirwa, Fred R.  
APPLICANT: Mahan, Suman M.  
TITLE OF INVENTION: Nucleic Acid Vaccines Against  
TITLE OF INVENTION: Rickettsial Diseases and Methods of Use  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Salivanchik & Salivanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/733,230  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Whiclock, Ted W.



```

Db      175 GATTCAAAAAATCTCAACCGTATTGGCTAAAAAAAGATTGGAGCGCTTAAACA 234
      80 LysThrAspAlaSerValGlyLeuSerAsnProSerAsnPheThr---IleProTyrThr 98
      235 CCATAGATTTCTAGC-----AATCTAATTTCTAAATTTTACGAAAAAGACATTTCT 288
      99 AlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrThrPheAla 117
      289 TTCAGATATGAAACAAATCCGTTTATAGTTTCGCTGAGCAATGGTACTCAATG--- 345
      118 GlnGlyThrArgValGlnIleGlnGlySerTyrGlnGlnPheAspValIleAsnProGly 137
      346 AATGACCAAGAGATAGAGTTCGAGATATCCATGAACTTTGATGTAAAAAACCAGT 405
      138 GlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGlnMelGlyAsn 157
      406 GGCAACTATAAACACGACACACATGTACTGTGCTTA-----GATACAGCACACAA 459
      158 SerPheThrProLysGlnLysValSerAsnSerIlePheHisThrValMetArgAsnAsp 177
      460 AATAGCACTAATGGCGCAGATTAATCTACTCTGT-----ATGTAATAAACAA 510
      178 GlyLeuSerIleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeu 197
      511 AATTAACAAATATATCATTAATGTTAAATGCGGTATGATGATCATGCTTGATGAGATA 570
      198 SerIleSerProTyrIleCysGlyGlyValAsnValAspAlaIleGlnPheAspVal 217
      571 CCAGTTTCCATATGATATGTCAGAGTATGGCAGCTGATAGTCTAGTAAATTAATCT 630
      218 LeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsn 237
      631 ACAATCTTAATTTCTTATTCAGAAAAGCTAGCATAAGTACTCAATTAATTTCAA 690
      238 IleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeu 257
      691 GCTTCTATCTTATGCGTGACATTTCCATAGAGTTATAGGTAATGAATTTAAA----- 744
      258 AsnValGlnHisValAlaGlnLeuAlaSerIlePheProLysIleThrSer----- 273
      745 -----GATATGCTACTACTTAAATAATTTACTTCAAAAAACAGAAATA 786
      274 -----AlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGlnIleGly 288
      787 TCTAATCTGCGCTTGCAATCAGCAACCTTATGTTGTCACTTTGGTATGAATTTGCA 846
      289 AlaArgLeuThrPhe 293
      847 GGAAGTTTGTATTT 861

RESULT 6
US-09-660-587-5
; Sequence 5, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 5
; LENGTH: 840
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; NAME/KEY: mat_peptide

```

```

; OTHER INFORMATION: nucleic acid sequence of p28-6
US-09-660-587-5

Alignment Scores:
Pred. No.: 6,74e-46 Length: 840
Score: 446.50 Matches: 105
Percent Similarity: 55.82% Conservative: 58
Best Local Similarity: 35.96% Mismatches: 110
Query Match: 29.85% Indels: 19
DB: 4 Gaps: 10

US-10-062-624-40 (1-293) x US-09-660-587-5 (1-840)
      6 LysPheThrIleIleAsnThrValLeuValCysLeuLeu---SerLeuProAsnIleSer 24
      10 AAAAAAATTTCTTAATCAACTGTCATTAATGCAATTAATGATGATGCTTCAAGCATATCT 69
      25 SerSerLysAlaIleAsnAsnAsnAlaLysLysTyrTyrGlyLeuTyrIleSerGlyGln 44
      70 TTTTCTGATCTATTAACAAGACATTAACCTGTAGCTTC-----TACATCAGTGAATA 123
      45 TyrLysProSerValSerValPheSerAsnPheSerValLysGlnThrAsnValIleThr 64
      124 TATGTACCAAGTGTTCACATTTGTGCTTTCTCACCTAAAGAAAGAAAGAACTCAACT 183
      65 LysAsnLeuIleAlaLeuValLysAspValAspSerIleGlnThrLysThrAspAlaSer 84
      184 GTTGAGAGTTTGGATTAATTAACATGATTTGGAATGA-----GTTACA 225
      85 ValGlyIleSerAsnProSerAsn---PheThrIlePro---TyrThrAlaValPheGln 102
      226 ATATCTAATCTTTCTCCAGAAATATATATTCACATTCATAATTTTCGTTAAATACGA 285
      103 AspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGlnGlyThrArg 121
      286 AACAAACCATTTCTTAGGTTTGCAGAGACTATGTTATTCAAATGCGT---GGCCCAAGA 342
      122 ValGlnIleGlnGlySerTyrGlnGlnPheAspValLysAsnProGlyGlyTyrThrLeu 141
      343 ATGAACCTTGAAAGTTCGTGACAGACATTCATGTAATAATTAATTAATTAATTAATG 402
      142 SerAspAlaTyrArgTyrPheAlaLeuAlaArgGlnMetLysGlyAsnSerPheThrPro 161
      403 AAGGCGCACACACATATCTGTGCTTATTCATCAT-----AGTTCAAGCAACAAGC 453
      162 LysGlnLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIle 181
      454 ATGTCTCCGCAAGTAACAAAATTT-----GTTTCTTAAATAATGAAGGTTAATTGAC 507
      182 IleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSerPro 201
      508 TTATCATTTATGATAAATGCAATGCTATGACATAATTAATGAAGAAAGCTTTTTCACCT 567
      202 TyrIleCysGlyGlyAlaGlyValAspAlaIleGlnPhePheAspValLeuHisIleLys 221
      568 TATATTTGTGAGAGTGTGTGACATGCTGTTTCCATGTTTGAAGCATTAATACCTTAA 627
      628 ATTTCTTACCAAGAAACTAGATAGTTAGTTATGATTAAGTTACAGAAAGCTGTTT 687
      222 PheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPhe 241
      628 ATTTCTTACCAAGAAACTAGATAGTTAGTTATGATTAAGTTACAGAAAGCTGTTT 687
      242 AlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGlnHis 261
      688 ATCGGTGACACTTTCACAGAGTCATAGTAAATTAATTAAGACATCCCTGTATAG--- 744
      262 ValAlaGlnLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIleGly 281
      745 GTTCTAGTGATTAATTTCTCCAGAAACCAATTTGCATATGTAACATTAATGTGTGT 804
      282 TyrPheGlyGlyGlnIleGlyValaArgLeuThrPhe 293
      805 CACTTGGCATAGAACTTGAAGAAATTAATTAATTTCACTTC 840

```

```

RESULT 7
/ Sequence 5, Application US/09261358A
/ Patent No. 6403780
/ GENERAL INFORMATION:
/ APPLICANT: Walker, David H.
/ APPLICANT: McBride, Jere W.
/ APPLICANT: Yu, Xue-jie
/ TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
/ TITLE OF INVENTION: Genes of Escherichia canis and Uses Thereof
/ FILE REFERENCE: D6152CIP
/ CURRENT APPLICATION NUMBER: US/09/261,358A
/ PRIOR FILING DATE: 1999-03-03
/ PRIOR APPLICATION NUMBER: 09/201,458
/ PRIOR FILING DATE: 1998-11-30
/ NUMBER OF SEQ ID NOS: 33
/ SEQ ID NO 5
/ LENGTH: 840
/ TYPE: DNA
/ ORGANISM: Escherichia canis
/ FEATURE:
/ NAME/KEY: mat peptide
/ OTHER INFORMATION: nucleic acid sequence of EC28SA3
US-09-261-358A-5

Alignment Scores:
Pred. No.: 6,74e-46 Length: 840
Score: 446.50 Matches: 105
Percent Similarity: 55.82% Conservative: 58
Best Local Similarity: 35.96% Mismatches: 110
Query Match: 29.85% Indels: 19
DB: 4 Gaps: 10

US-10-062-624-40 (1-293) x US-09-261-358A-5 (1-840)

Qy 6 LysPhehrllellelsnhrhvalleuvalCysleuLeu---SerLeuProsnlleSer 24
Db 10 AAAAAAATTCTTAAACAACGCTGATTAAGTCATTAAAGTACTATGCTCCAGAGATATCT 69

Qy 25 SerSerLyalaalleasnansenalalyslyrTyrglyLeuTyrlleSerglyGln 44
Db 70 TTTTTCGATCTATACAGACGATACACTGCTAGCTTC-----TACATCAGGGAAA 123

Qy 45 TyrlsPProSerValserValPheSerAenPheSerVallysglulThraenValilleThr 64
Db 124 TATGTACTCAAGTCTTTCACATTTGGTGTCTTCTCAGCTAAAGAAAGAACTCAACT 183

Qy 65 LysAenleuilealeuLysLyasPValasPserilleglulThrlyeThraPsalSer 84
Db 184 GTTCGAGCTTTTGGATTAAACATGATTCGGAATGCA-----GCTACA 225

Qy 85 ValgllylleserAenProSerAen--PheThrilePro--TyrlThralaValPheGln 102
Db 226 ATATCTAACTCTTCTCCAGAAATATATATTCACAGTCAAAATATATTCGTTAAATACGAA 285

Qy 103 AspaenSer---ValasnPheAenglyThrillegllyrThrPhealaglulThraArg 121
Db 286 AACAAACCATCTTAGGGTTTCGAGGAGCTATTTGGTATTCATCAGGGT--GGCCCAAGA 342

Qy 122 ValglulileglulyseryrTyrglulPheasPVallyasPProgllyTyrlThreU 141
Db 343 ATGAACTTGAAGTTCGTGTCAGAGATTCGATGCGTGAAGAAATACAGAAATATTTATAG 402

Qy 142 SerasPAlarYrArgrYrPhealaleualarglumelYsglyAsenSerPheThrPro 161
Db 403 AACGCCACACAGATACGTGCTTATCTCATCAT-----AGTTCGCAACAGC 453

Qy 162 LysglulysValSerAenSerillePheHesthrValMeArGAsenAspGlyLeuSerille 181
Db 454 ATGTCCTCCGCAAGTACCAATTT-----GTTTCTTAAATAAGAGGGTTAATTGAC 507

Qy 182 lleserValillevalenValCysTyraPpPheSerleuAsnAenleuSerilleSerPro 201
Db 507

```

```

Db      508 TTATCATTTATGATAAAGTCATGCATATGACATAATATTGAAGAATGCCCTTTTCAACT 567
Qy      202 TrrtlleCgGglYglYAlAglyValnepralalleGluPherphesapValLeuhlsileys 221
Db      568 TATATTTTGCGAGGGGTGGTAGCTATGTTGTTCCATGTTTTGAACCTTAATAATCCCTAAA 627
Qy      222 PhealrYrgInSerIseLeuglYllealaYrSerLeuProSeasnlleSereLeuPhe 241
Db      628 ATTCTTTNCCAAGAAATCTAGATTTAGGTTATAGTATTAAGTTACAAGCCTCTGTTTTT 687
Qy      242 AlaSerLeuYrTyRHisLyVaImecGlYaenGlnPhelYsaLnLubsnValGlnHIs 261
Db      668 ATCGGTGACACTTTCACAGAGCTCATAGTAAATGAATTTAGAACATCCCTGCTATG-- 744
Qy      262 ValAlaGluLeuAlaSerIleProLySlleThrSerAlaValAlaThrLeuAsnilleGly 281
Db      745 GTTCTCTAGTGATCAAAATCTTCCAGAAAAACAATTTGCATATAGTAACTAAATGTGTGT 804
Qy      282 TyrPheGlgYglYglUlleGlyAlaArgLeuThrPhe 293
Db      805 CACTTTGGCATGAACTTGGAGAAGATTTAACTTC 840

RESULT 8
US-09-660-587-1
; Sequence 1, Application US/09660587
; Patent No. 6392023
GENERAL INFORMATION:
APPLICANT: Walker, David H.
APPLICANT: McBride, Jere W.
APPLICANT: Xu, Xue-Jie
TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
FILE REFERENCE: D6152CIP2
CURRENT APPLICATION NUMBER: US/09/660,587
CURRENT FILING DATE: 2000-09-12
PRIOR FILING DATE: 09/261,358
NUMBER OF SEQ ID NOS: 46
SEQ ID NO 1
LENGTH: 1607
TYPE: DNA
ORGANISM: Ehrlichia canis
FEATURE:
OTHER INFORMATION: nucleic acid sequence of E. canis p28-7
US-09-660-587-1

Alignment Scores:
Pred. No.: 4,32e+45 Length: 1607
Score: 443.50 Matches: 106
Percent Similarity: 54.18% Conservative: 56
Best Local Similarity: 35.45% Mismatches: 110
Query Match: 29.65% Indels: 27
DB: 4 Gaps: 8

US-10-062-624-40 (1-293) x US-09-660-587-1 (1-1607)
Qy      2 AenAnLuVleuLYperHerTrIlleAenThrValLeuValCySteLeuLeu---SerLeu 20
Db      143 AATATGAATTGCAAAAAAATTTCTTATMACACTGCATTATATCATTAAGTACTAT 202
Qy      21 ProAnllleSeSerSerLySaLaIleAenAenAnAlaLyseLyTyRtyGlyLeuTYr 40
Db      203 CCACAGCATCTTTTCTTCATCTATTA---CAAGATGATMACTGGGTGTAACCTTAT 255
Qy      41 llesGrlYglNYrLYsProSeValSerValPheSeasnPheSeVallysgIuThr 60
Db      260 ATTATGGAAGTAATGACCAAGTGTCTCACATTTTGGTAGCTTCTCAGCTAAAGAA 319
Qy      61 AenValIleThrLYsaLnLeuIleAlaLeuLYsLYsaSPValAlaSerIleGluThrlys 80
Db      320 AGCAAATCAACTGTTGGAGATTTTGGATTAAACATGATTTGGATGGA----- 367
Qy      81 ThrApAlaSerValGIvlyleSerAenProSerAenPheThrIlePro--TYrThAla 99

```

```

Db      368 -----AGTCCAACTTAAAGATTAACACGCTGACTTACTGTTCCAACTATTCGTTCC 421
Qy      100 ValPheGlnAAspAAsenSer---ValAsnPhenGlnGlyThrIleGlyTyrThrPheAlaGlu 118
Db      422 AGATACGAGAACAACTTCTTACGGTTTGACAGAGCTATCGGTACTACCAATGGGT--- 478
Qy      119 GLyThrArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGlyGly 138
Db      479 GGGCCCAAGATGATTCGAATATCTTATGAGCATTCGACGTAAAAAGTCTTAATATC 538
Qy      139 TyrThrLysSerAspAlaTyrTrpTyrPheAlaLeuAlaArgGlu-----Met 154
Db      539 AATTATCAAAATGACGCGACAGTACGCGCTATCTCATCACACATCGGCAGCCATG 598
Qy      155 LysGlyAAsenSerPheThrProLysGlyLysValSerAAsenSerIlePheIsthValMet 174
Db      599 GAAGCTGATTAATTT-----GTCCTCTTA 622
Qy      175 ArgAsnAAspGlyLysSerIleIleSerValIleValAsnValCysTyrAspPheSerLeu 194
Db      623 AAAAAAGAGGTTAATTGACATATCACTGCATTAATATGATGATTAATAAT 682
Qy      195 AsnAAsenLysSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhe 214
Db      683 GACAAAGTACCTGTTCTTCCTTATATATGCGCAGGATTTGTAAGTATTTCTATG 742
Qy      215 PheAspValLysIleGlyIleGlyPheAlaTyrGlnSerIleValAlaTyrSerLeu 234
Db      743 TTTGAAGTACAACTGCTTAATTTCTTACCAAGAAACCTGGGATTAAGTACTCTAT 802
Qy      245 ProSerAAsnIleSerLeuPheAlaSerLeuTyrTyrIleLysValMetGlyAsnGlnPhe 254
Db      803 AATCGGAAACCTCTGTTTTCATCGTGCGCATTTCCACAGATATAGTATATAGATT 862
Qy      255 LysAAsenLysAAsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAla 274
Db      863 AGAGTATTTCTGCAATAGTACCTAGTAACTCACTCACTAATAGTGACCAATTTCCA 922
Qy      275 ValAlaThrLeuAAsnIleGlyTyrPheGlyGlyIleGlyValAlaArgLeuThrPhe 293
Db      923 ACAGTAACTAAATGATGTGTCACCTTGTTAAGAACTTGAGGAAATTTAACTTC 979

RESULT 9
US-09-261-358A-1
; Sequence 1, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 1
; LENGTH: 1607
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: nucleic acid sequence of Eca28-1
US-09-261-358A-1

Alignment Scores:
Pred. No.: 4,328-45 Length: 1607
Score: 443.50 Matches: 106
Percent Similarity: 54.18% Conservative: 56
Best Local Similarity: 35.45% Mismatches: 110
Query Match: 29.65% Indels: 27

```

```

DB:      4 Gaps:      8
US-10-062-624-40 (1-293) x US-09-261-358A-1 (1-1607)
Qy      2 AsnAAsnLysLeuLysPheThrIleIleAsnThrValLeuValCysLeuLeu---SerLeu 20
Db      143 AATATGATTTCAAAAAAATCTTATTAACAACATCCATTAATATCATTAATGACTCTAT 202
Qy      21 ProAAsnIleSerSerSerLysAlaIleAsnAAsnAAsnAlaLysLysTyrTyrGlyLeuTyr 40
Db      203 CCAAGCATATCTTTTCTGATACATATA---CAAGATGTAACATGGGTGTAACCTTCTAT 259
Qy      41 IleSerGlyGlnTyrLysProSerValSerValPheSerAAsnPheSerValIleGluThr 60
Db      260 ATTAGTGAAGAAGTATGTCACAGTCTCTCAATTTTGGTACTTCTCAGCTTAAGAAAGAA 319
Qy      61 AsnValIleThrLysAAsnLeuIleAlaLeuLysLysAAspValAAspSerIleGluThrLys 80
Db      320 AGCAAAATCACTGTTGGAGTTTGGATTAATAACATGATGGATGGA----- 367
Qy      81 ThrAspAlaSerValGlyIleSerAAsnProSerAAsnPheThrIlePro---TyrThrAla 99
Db      368 -----AGTCCAACTTAAAGATTAACACGCTGACTTACTGTTCCAACTATTCGTTCC 421
Qy      100 ValPheGlnAAspAAsenSer---ValAsnPhenGlnGlyThrIleGlyTyrThrPheAlaGlu 118
Db      422 AGATACGAGAACAACTTCTTACGGTTTGACAGAGCTATCGGTACTACCAATGGGT--- 478
Qy      119 GLyThrArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGlyGly 138
Db      479 GGGCCCAAGATGATTCGAATATCTTATGAGCATTCGACGTAAAAAGTCTTAATATC 538
Qy      139 TyrThrLysSerAAspAlaTyrTrpTyrPheAlaLeuAlaArgGlu-----Met 154
Db      539 AATTATCAAAATGACGCGACAGTACGCGCTATCTCATCACACATCGGCAGCCATG 598
Qy      155 LysGlyAAsenSerPheThrProLysGlyLysValSerAAsenSerIlePheIsthValMet 174
Db      599 GAAGCTGATTAATTT-----GTCCTCTTA 622
Qy      175 ArgAAsnAAspGlyLysSerIleIleSerValIleValAsnValCysTyrAspPheSerLeu 194
Db      623 AAAAAAGAGGTTAATTGACATATCACTGCATTAATATGATGATTAATAAT 682
Qy      195 AsnAAsenLysSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhe 214
Db      683 GACAAAGTACCTGTTCTTCCTTATATATGCGCAGTATTTGTAAGTATTTCTATG 742
Qy      215 PheAspValLysIleGlyIleGlyPheAlaTyrGlnSerIleGlyIleAlaTyrSerLeu 234
Db      743 TTTGAAGTACAACTGCTTAATTTCTTACCAAGAAACCTGGGATTAAGTACTCTAT 802
Qy      245 ProSerAAsnIleSerLeuPheAlaSerLeuTyrTyrIleLysValMetGlyAsnGlnPhe 254
Db      803 AATCGGAAACCTCTGTTTTCATCGTGCGCATTTCCACAGATATAGTATATAGATT 862
Qy      255 LysAAsenLysAAsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAla 274
Db      863 AGAGTATTTCTGCAATAGTACCTAGTAACTCACTCACTAATAGTGACCAATTTCCA 922
Qy      275 ValAlaThrLeuAAsnIleGlyTyrPheGlyGlyIleGlyValAlaArgLeuThrPhe 293
Db      923 ACAGTAACTAAATGATGTGTCACCTTGTTAAGAACTTGAGGAAATTTAACTTC 979

RESULT 10
US-09-648-520B-47
; Sequence 47, Application US/09648520B
; Patent No. 6432649
; GENERAL INFORMATION:
; APPLICANT: Stich, Roger W.
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Methods for Detecting Ehrlichia Canis and Ehrlichia Chaffeensis 1
; TITLE OF INVENTION: Vertebrate and Invertebrate Hosts

```

```
FILE REFERENCE: 22727/04069
CURRENT APPLICATION NUMBER: US/09/648,520E
CURRENT FILING DATE: 2000-08-25
NUMBER OF SEQ ID NOS: 51
SOFTWARE: Patent version 3.1
SEQ ID NO 47
LENGTH: 1607
TYPE: DNA
ORGANISM: JAKE strain, Ehrlichia canis p30
US-09-648-520E-47

Alignment Scores:
Pred. No.: 4,32e-45 Length: 1607
Score: 443.50 Matches: 106
Percent Similarity: 54.18% Conservative: 56
Best Local Similarity: 35.45% Mismatches: 110
Query Match: 29.65% Indels: 27
DB: 4 Gaps: 8

US-10-062-624-40 (1-293) x US-09-648-520E-47 (1-1607)

QY 2 AasnAnlySLeuYpHeThrIleIleAnThrValLeuValCysLeuLeu---SerLeu 20
DB 143 AATATGAATTGCAAAAAAATTTCTTATACCACTGCATTATATCATATGACTCTAT 202
QY 21 ProAnIleSerSerSerysAlaIleAenAenAlaIlyLeuTyTyGlyLeuTy 40
DB 203 CCAAGCATATCTTTTCTGATCTATTA---CAAGATGTAACATGGGCTGACTCTAT 259
QY 41 IleserGlyGlnTyTyLysProSerValSerValPheSerAenPheSerValIysGluThr 60
DB 260 ATTATGGAAGAAATAGTACCAAGTGTCTCACATTTTGGAGCTTCTCAGCTTAAGAGAA 319
QY 61 AenValIleThrLysAenLeuIleAlaLeuIlySlyAspValAspSerIleGluThrLys 80
DB 330 AGCAAAATCAACTGTGAGATTTTGCATTAACATGATGGAGATGA----- 367
QY 81 ThrAapAlaSerValIglyIleSerAenProSerAenPheThrIlePro---TyThrAla 99
DB 368 -----AGTCCAAATCTTAAAGATTAACACGCTGACTTACTCTTCAAACTATTCGTT 421
QY 100 ValPheGlnAspAenSer---ValAenPheAenGlyThrIleGlyTyThrPheAlaGlu 118
DB 422 AGATACGAGAACATTCATTCATAGGCTTTCAGAGGCTATCGATGATCTCAATGGGT--- 478
QY 119 GlyThrArgValGluIleGluGlySerTyTyGluGluPheAspValIlyAenProGlyGly 138
DB 479 GGCACCAAGATGAATTCGAAATATCTTAAGAAGCATTCGACGTAAAGTCTAATATC 538
QY 139 TyrThrLeuSerAspAlaTyArgTyRheAlaLeuAlaArgLys-----Met 154
DB 539 AATTATCAAAATGACGCGCACAGGTCGCTCTATCTCATCACACATCGGACCATG 598
QY 155 LysGlyAenSerPheThrProLysGluIySlySerAenSerIlePheHisThrValMet 174
DB 599 GAAGCTGATTAATTT-----GTTCTTCTTA 622
QY 175 ArgAenAspGlyLeuSerIleIleSerValIleValAenValCysTyThrAspPheSerLeu 194
DB 623 AAAAAGCAAGGCTTATGACATATCATCTTGCAATTAAGATGATTAATTAAT 682
QY 195 AenAenLeuSerIleSerProTyTyIleCysGlyGlyAlaGlyValAspAlaIleGluPhe 214
DB 683 GACAAAGTACCTGTTCTTCTTATATGCGCAGGATATGAGTATGATTTCTATG 742
QY 215 PheAspValLeuHisIleLysPheAlaTyArgIleSerValLeuGlyIleAlaTySerLeu 234
DB 743 TTGAAAGCTACAAAGTCTTAATAATTTCTTCCACAGGAAATCGGGCATTTGATCTAT 802
QY 235 ProSerAenIleSerLeuPheAlaSerLeuTyTyTyHisIleValMetGlyAenGlnPhe 254
DB 803 AATCCGAAACCTGTTTTCATCGGTGGCATTTTCCACAGATCATAGTAAATGATTT 862
```

```
QY 255 LysAenLeuAenValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAla 274
DB 863 AGAATATTCCTGCAATAGTACCTGATCACTCACTCACTCACTCACTCACTCACTCACT 922
QY 275 ValAlaThrLeuAenIleGlyTyRheGlyGlyGluIleGlyAlaArgLeuThrPhe 293
DB 923 ACAGTAACACTTAATGTGTGTCACCTTGTGTTAGACTGTGAGAGATTTTACTTC 979

RESULT 11
US-09-201-458-1
Sequence 1, Application US/09201458A
Patent No. 6458942
GENERAL INFORMATION:
APPLICANT: Walker, David H.
APPLICANT: McBride, Jere W.
APPLICANT: Yu, Xue-Jie
TITLE OF INVENTION: 28-KDa Immunoreactive Protein Gene of Ehrlichia
FILE REFERENCE: D6152
CURRENT APPLICATION NUMBER: US/09/201,458A
CURRENT FILING DATE: 1998-11-30
NUMBER OF SEQ ID NOS: 21
SEQ ID NO 1
LENGTH: 1607
TYPE: DNA
ORGANISM: Ehrlichia canis
FEATURE:
OTHER INFORMATION: nucleic acid sequence of a gene encoding a 30 kDa
immunoreactive protein of Ehrlichia canis
US-09-201-458-1

Alignment Scores:
Pred. No.: 4,32e-45 Length: 1607
Score: 443.50 Matches: 106
Percent Similarity: 54.18% Conservative: 56
Best Local Similarity: 35.45% Mismatches: 110
Query Match: 29.65% Indels: 27
DB: 4 Gaps: 8

US-10-062-624-40 (1-293) x US-09-201-458-1 (1-1607)

QY 2 AasnAnlySLeuYpHeThrIleIleAnThrValLeuValCysLeuLeu---SerLeu 20
DB 143 AATATGAATTGCAAAAAAATTTCTTATACCACTGCATTATATCATATGACTCTAT 202
QY 21 ProAnIleSerSerSerysAlaIleAenAenAlaIlyLeuTyTyGlyLeuTy 40
DB 203 CCAAGCATATCTTTTCTGATCTATTA---CAAGATGTAACATGGGCTGACTCTAT 259
QY 41 IleserGlyGlnTyTyLysProSerValSerValPheSerAenPheSerValIysGluThr 60
DB 260 ATTATGGAAGAAATAGTACCAAGTGTCTCACATTTTGGAGCTTCTCAGCTTAAGAGAA 319
QY 61 AenValIleThrLysAenLeuIleAlaLeuIlySlyAspValAspSerIleGluThrLys 80
DB 320 AGCAAAATCAACTGTGAGATTTTGCATTAACATGATGGAGATGA----- 367
QY 81 ThrAapAlaSerValIglyIleSerAenProSerAenPheThrIlePro---TyThrAla 99
DB 368 -----AGTCCAAATCTTAAAGATTAACACGCTGACTTACTCTTCAAACTATTCGTT 421
QY 100 ValPheGlnAspAenSer---ValAenPheAenGlyThrIleGlyTyThrPheAlaGlu 118
DB 422 AGATACGAGAACATTCATTCATAGGCTTTCAGAGGCTATCGATGATCTCAATGGGT--- 478
QY 119 GlyThrArgValGluIleGluGlySerTyTyGluGluPheAspValIlyAenProGlyGly 138
DB 479 GGCACCAAGATGAATTCGAAATATCTTAAGAAGCATTCGACGTAAAGTCTAATATC 538
QY 139 TyrThrLeuSerAspAlaTyArgTyRheAlaLeuAlaArgLys-----Met 154
DB 539 AATTATCAAAATGACGCGCACAGGTCGCTCTATCTCATCACACATCGGACCATG 598
```



ORGANISM: Ehrlichia canis  
 FEATURE:  
 NAME/KEY: mat\_peptide  
 OTHER INFORMATION: nucleic acid sequence of p28-5  
 US-09-660-587-3

## Alignment Scores:

Pred. No.:	3, 28e-45	Length:	849
Score:	441.00	Matches:	109
Percent Similarity:	52.54%	Conservative:	46
Best Local Similarity:	36.95%	Mismatches:	126
Query Match:	29.48%	Indels:	14
DB:	4	Gaps:	7

US-10-062-624-40 (1-293) x US-09-660-587-3 (1-849)

```

QY      1 MetAsnAenLYLeuLYsPheThrIleIleAsnThrValLeuValCysLeuLeuSerLeu 20
DB      1 ATGAATTGTAAAGTTTACACA--ATAAGTGCATTGATATCATCATATATCTCCCA 57
QY      21 ProAenIleSerSerSerSerValaIleAsnAenAenAlaLYsTYrTYrGlyLeuTYr 40
DB      58 CCTAATGCTCTACTACTTAACCCAGTATATGTTAGTACAGT--ATGATAGTAAATTTTAC 114
QY      41 IleSerGlyGlnTYrLYsProSerValSerValPheSerAsnPheSerValLYsGlyThr 60
DB      115 AATTCAGAAAGTACATGCCAAGTGTCTCTCATTTTGGAAATTTTTCAGCTGAAAGAG 174
QY      61 AsnValIleThrLYsAsnLeuIleAlaLeuLYsLYsAspValAspSerIleGlyThrLYs 80
DB      175 AAAAAAAGACAACCTGATATATGCTTAAAGAAAC-----TGCGCA 219
QY      81 ThrAspAlaSerValGlyLYsSerAsnProSerAsnPheThrIle---ProTYrThra 99
DB      220 GGAGATGCAATATCTAGTCAAGTCCAGATATATTTTACCATTCGAAATTTACTCATTC 279
QY      100 ValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTYrThrPheAlaGlu 118
DB      280 AGTATGCAAGCAACAAGTTTATAGGCTTTGCGATGATGATTTGTTACTCGATAGGC--- 336
QY      119 GlyThrArgValGluIleGlySerTYrGlyGluPheAspValLYsAsnProGlyLYs 138
DB      337 AGTCCAAAGATAGAGTGTATGATGCTTATGAAGCATTTGATGAAAAATCCAGTGTAT 396
QY      139 TYrThrLeuSerAspAlaTYrArgTYrPheAlaLeuAlaArgLYsMetLYsGlyAsnSer 158
DB      397 AATTACAAAACGGGTCTTACAGATATGCTTATCTCATCATCAAGATATCGGATGAT 456
QY      159 PheThrProLYsGlyLYsValSerAsnSerIlePheIleThrValMetArgAsnAspGly 178
DB      457 -----GACATGACTAGTGCACACTGACCAATTTGTATATTATTAATGAAGCA 504
QY      179 LeuSerIleIleSerValIleValAsnValCysTYrAspPheSerLeuAsnLeuSer 198
DB      505 TTAATTAACATATCATTTATGACAAACATATGTTAAGAAACAGCAAGCAAAATATACCT 564
QY      199 IleSerProTYrIleCysGlyLYsAlaGlyValAspAlaIleGluPhePheAspValLeu 218
DB      565 CTTCTCTTACATATGATGCAAGTATGCTGATGATTTATTCACATGTTTGAAGTAC 624
QY      219 HisIleLYsPheAlaTYrGlnSerLYsLeuGlyIleAlaTYrSerLeuProSerAsnIle 238
DB      625 CATCCAAAATTTCTTATCAAGAAAGCTAGGTTGCTTACTGCTGTAAGTCAAGTGT 684
QY      239 SerLeuPheAlaSerLeuTYrTYrHisLYsValMetGlyAsnGlnPheLYsAsnLeuAsn 258
DB      685 TCGGTTCTTTTGGTATATATTTTCAATAAATATTAATTAATGAAGTAAATGTTCCA 744
QY      259 ValGlnHisValaIleGluLeuAlaSerIleProLYsIleThrSerValaIleThrIle 278
DB      745 GCCATGGTACTTATTAATCAAGCAAGATATGAGCAACAGTTTCCAAAGTAAACATTA 804
QY      279 AsnIleGlyTYrPheGlyGlyIleGlyAlaArgLeuThrPhe 293
  
```

DB 805 AATGATGCTACTTGTGATTAAGAACTTGATATGATGTTCAACTTC 849

## RESULT 14

US-09-261-358A-3  
 ; Sequence 3, Application US/09261358A

; Patent No. 6403780

; GENERAL INFORMATION:

; APPLICANT: Walker, David H.

; APPLICANT: McBride, Jere W.

; APPLICANT: Yu, Xue-Die

; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein

; FILE REFERENCE: D6152CIP

; CURRENT APPLICATION NUMBER: US/09/261,358A

; PRIOR APPLICATION NUMBER: 09/201,458

; PRIORITY FILING DATE: 1998-11-30

; NUMBER OF SEQ ID NOS: 33

; SEQ ID NO 3

; LENGTH: 849

; TYPE: DNA

; ORGANISM: Ehrlichia canis

; FEATURE:

; NAME/KEY: mat\_peptide

; OTHER INFORMATION: nucleic acid sequence of Eca28SA2

US-09-261-358A-3

## Alignment Scores:

Pred. No.:	3, 28e-45	Length:	849
Score:	441.00	Matches:	109
Percent Similarity:	52.54%	Conservative:	46
Best Local Similarity:	36.95%	Mismatches:	126
Query Match:	29.48%	Indels:	14
DB:	4	Gaps:	7

US-10-062-624-40 (1-293) x US-09-261-358A-3 (1-849)

```

QY      1 MetAsnAenLYLeuLYsPheThrIleIleAsnThrValLeuValCysLeuLeuSerLeu 20
DB      1 ATGAATTGTAAAGTTTACACA--ATAAGTGCATTGATATCATCATATATCTCCCA 57
QY      21 ProAenIleSerSerSerSerValaIleAsnAenAenAlaLYsTYrTYrGlyLeuTYr 40
DB      58 CCTAATGCTACTACTTAACCCAGTATATGTTAGTACAGT--ATGATAGTAAATTTTAC 114
QY      41 IleSerGlyGlnTYrLYsProSerValSerValPheSerAsnPheSerValLYsGlyThr 60
DB      115 AATTCAGAAAGTACATGCCAAGTGTCTCTCATTTTGGAAATTTTTCAGCTGAAAGAG 174
QY      61 AsnValIleThrLYsAsnLeuIleAlaLeuLYsLYsAspValAspSerIleGlyThrLYs 80
DB      175 AAAAAAAGACAACCTGATATATGCTTAAAGAAAC-----TGCGCA 219
QY      81 ThrAspAlaSerValGlyLYsSerAsnProSerAsnPheThrIle---ProTYrThra 99
DB      220 GGAGATGCAATATCTAGTCAAGTCCAGATATATTTTACCATTCGAAATTTACTCATTC 279
QY      100 ValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTYrThrPheAlaGlu 118
DB      280 AAGTATGCAAGCAACAAGTTTATAGGCTTTGCGATGATGATTTGTTACTCGATAGGC--- 336
QY      119 GlyThrArgValGluIleGlySerTYrGlyGluPheAspValLYsAsnProGlyLYs 138
DB      337 AGTCCAAAGATAGAGTGTATGATGCTTATGAAGCATTTGATGAAAAATCCAGTGTAT 396
QY      139 TYrThrLeuSerAspAlaTYrArgTYrPheAlaLeuAlaArgLYsMetLYsGlyAsnSer 158
DB      397 AATTACAAAACGGGTCTTACAGATATGCTTATCTCATCATCAAGATATCGGATGAT 456
QY      159 PheThrProLYsGlyLYsValSerAsnSerIlePheIleThrValMetArgAsnAspGly 178
DB      457 -----GACATGACTAGTGCACACTGACCAATTTGTATATTATTAATGAAGCA 504
  
```





US-08-953-326-8

## Alignment Scores:

Pred. No.:	1,81e-43	Length:	861
Score:	427.00	Matches:	108
Percent Similarity:	56.12%	Conservative:	57
Best Local Similarity:	36.73%	Mismatches:	109
Query Match:	28.54%	Indels:	20
DB:	4	Gaps:	12

US-10-062-624-40 (1-293) x US-08-953-326-8 (1-861)

```

QY 6 LysPheThrIleIleAsnThrValIleuValCysLeuLeuSer---LeuProAsnIleSer 24
DB 15 AAATTT---TTTATACAACTGCATTACATTACGATGCTCTTCTTACCTGATATCA 71
QY 25 SerSerLysAlaIle---AsnAsnAsnAlaLysLysTyTyTyGlyLeuTyTlleSerGly 43
DB 72 CTTTCTGATCCAGTACAGATGACAAACATTAGTGGTAATTTC-----TACATCAGTGA 125
QY 44 GlnTyTlrsProSerValSerValPheSerAsnPheSerValValGluThrAsnValIle 63
DB 126 AAGTATATGCGCAAGCGCTTCGATTTTTCGATTTTTCGCCAAGAGAAAGAAATACA 185
QY 64 ThrLysAsnLeuIleAlaLeuLysLysAspValAsp---SerIleGluThrLysThrAsp 82
DB 186 ACGATTGAGATTTTGGATATAGCAAGATTGGCATAGATGATATATCTAGAACCACT 245
QY 83 AlaSerValGlyIleSerAsnProSerAsnPheThrIlePro---TyThrAlaValPhe 101
DB 246 TTAACG-----GATATATTACCGCTTCACAAATTATTCATTATTAAGTAT 287
QY 102 GlnAspAsnSerValAsn---PheAsnGlyThrIleGlyTyThrPheAlaGluGlyThr 120
DB 288 GAAATTAATCTATTTTCAGATTCAGAGATTGCGAGCTATTGGTACTCATG---GATGCCCA 344
QY 121 ArgValGluIleGluGlySerTyTlrsGluIlePheAspValLysAsnProGlyGlyTyThr 140
DB 345 AGAATAGAGCTTGAAGTATCTTATAGACATTTCGATGTTAAATACAGATACAAATAT 404
QY 141 LeuSerAspAlaTyTlrsTyThrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThr 160
DB 405 AAGAAGCAAGCACATAGATATATATGCTGCTCC---CATCTTCGCGCACAGACAG 461
QY 161 ProLysGluLysValSerAsnSerIlePheIleThrValMetArgAsnAspGlyLeuSer 180
DB 462 ATAGATGTCAGCGCGCTGCTGCTTT-----CTAATTAATGAAGAGCTACTT 512
QY 181 IleIleSerValIleValAsnValCysTyTlrsPheSerLeuAsnAsnLeuSerIleSer 200
DB 513 GATTAATCATTTATGCTGAAGCATTTATGATGTAATGAAGAGCATACCTTTTCT 572
QY 201 ProTyTlrsGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIle 220
DB 573 CCTTATATATGTCAGGTATGCTATGATTGATTAATCATTCATGCTTAATATCT 632
QY 221 LysPheAlaTyTlrsGlySerLysLeuGlyIleAlaTyTlrsSerLeuProSerAsnIleSerLeu 240
DB 633 AAAAATTTCTTATACAGAAATTTAGGCTTAAGTACCTTAAGCCCAAGACCTTG 692
QY 241 PheAlaSerLeuTyTlrsIleLysValMetGlyAsnGlnPheLysAsnLeuAsnValGln 260
DB 693 TTTATTCGAGCATTTTCATTAAGGTATGAGAAAGCAATTTTAGAGATTTCTCTATG 752
QY 261 HisValAlaGluLeuAlaSerIleProLysIleThr---SerAlaValAlaThrLeuAsn 279
DB 753 ATACCTAGGAATCGACGCTTCAGAGAAAGCAATCCCTGCAATAGTAACTGAC 812
QY 280 IleglyTyTlrsGlyGlyValIleGlyAlaArgLeuThrPhe 293
DB 813 GTGTTCTACTTTGGCATAGACTTGGAGGAGGTTTAACCTT 854

```

RESULT 17

US-08-953-326-12

Sequence 12, Application US/08953326

Patent No. 6251872

## GENERAL INFORMATION:

APPLICANT: Bardet, Anthony F.  
 APPLICANT: Ganca, Roman R.  
 APPLICANT: McGuire, Travis C.  
 APPLICANT: Burridge, Michael J.  
 APPLICANT: Nyika, Aceme  
 APPLICANT: Rurangirwa, Fred R.

TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of

FILE REFERENCE: US-167C1

CURRENT APPLICATION NUMBER: US/08/953,326

EARLIER FILING DATE: 1997-10-17

EARLIER APPLICATION NUMBER: 08/953,326

EARLIER FILING DATE: 1997-10-17

NUMBER OF SEQ. ID NOS: 24

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 12

LENGTH: 864

TYPE: DNA

ORGANISM: Ehrlichia canis

US-08-953-326-12

## Alignment Scores:

Pred. No.:	2,41e-43	Length:	864
Score:	426.00 <td>Matches:</td> <td>110 </td>	Matches:	110
Percent Similarity:	51.57% <td>Conservative:</td> <td>54 </td>	Conservative:	54
Best Local Similarity:	34.59% <td>Mismatches:</td> <td>56 </td>	Mismatches:	56
Query Match:	28.48% <td>Indels:</td> <td>56 </td>	Indels:	56
DB:	4	Gaps:	13

US-10-062-624-40 (1-293) x US-08-953-326-12 (1-864)

```

QY 1 MetAsnAsnLysLeuLysPheThrIleIleAsnThrValIleuValCysLeuLeuSer--- 19
DB 3 ATGAATATATTAATAAACTTTTACAGTA-----ACTGATTAATTAATTTAACTTCTT 56
QY 20 -----LeuProAsnIleSerSerSerSerLysAlaIleAsnAsnAlaLysLysTyT 36
DB 57 ACACATTTTATACCTTTTATATAGTCCAGCAGCT-----GCCAGTACAAATT 101
QY 37 TyTlrsGlyTyTlrsGlyGlyGlyTyTlrsPheProSerValSerValPheSerAsnPheSer 56
DB 102 CACAACCTTCTCATTTAGTGAATAATATATGCAACAGCGTCACATTTTGAATTTTCA 161
QY 57 ValLysGluThrAsnValIleThrLysAsnLeuIleAlaLeuLysAspValAsp--- 75
DB 162 GCTTAAGAGAACAAAGTTTATCTTAAGGTATTAAGTGGCTTAAGTCAACGATTAACAT 221
QY 76 SerIleGluThrLysThrAspAla-----SerValGlyIleSerAsnProSerAsnPhe 93
DB 222 AATATTATTAACCAATATGATATAGACAGAAAGCTTAAGTTCATAAT----- 269
QY 94 ThrIleProTyTlrsAlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIle 112
DB 270 -----TATTCATTTAATTAACAAATAATCAATTCATTAAGTTTGCAGAGCTAT 320
QY 113 GlyTyTlrsPheAlaGluGlyTyTlrsArgValGluIleGluGlySerTyTlrsGluIlePheAsp 132
DB 321 GGTATTTAATAGGC---AATCAAGATTAACCTTAAGATATCAATGAATATTTAT 377
QY 133 ValLysAsnProGlyGlyTyTlrsLeuSerAspAlaTyTlrsTyTlrsPheAlaLeuAlaArg 152
DB 378 ACTAAAAACCAAGAAACAATATTTAAATATGACTCAAAATATGCGCTTATCTCAT 437
QY 153 Glu-----MetLysGlyAsnSer-----PheThrProLysGluLysVal 165
DB 438 GGAAGTCATATATGACGATGATGAATAATAGCGAGATTTGTACACTGCAAAAATGATTAAG 497

```

```

Qy 166 SerAsnSerIlePheHisIthrValMetArgAsnAspGlyLeuSerIleIleSerValIle 185
Db 498 -----TTTGTACTTCGAAAATAAGAGTTTACTGACGCTCATTTATG 542
Qy 186 ValAsnValCysTyrAspPheSerIleuAsnLeuSerIleSerProTyrIleCysGly 205
Db 543 TTAAGCGCATGTGATGACATACAACTGAATAATGCTTTTACCTTATATATGTGCA 602
Qy 206 GlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIleLeuPheAlaTyrGln 225
Db 603 GGTATGTGACTGATCTCATATCTATGTTGACAAACAACAAACAAATATATCTATCA 662
Qy 226 SerTyrLeuGlyIleAlaTyrSerIleuProSerAsnIleSerIlePheAlaSerLeuTyr 245
Db 663 GGAAGTTATGCTTTAACTATACATTAACCTCAAGAGTTCTGCTTTTGCAGGTGGCAC 722
Qy 246 TyrHisIlyValMetGlyValAsnGlnPheIys----- 255
Db 723 TTTCAATAGGTATAGTATGATTAAGGATTTCTCTACTATTAACCTGATGATCA 782
Qy 256 AsnLeuAsnValGlnHisIValAlaGluLeuAlaSerIleProIyalleIthrSerAlaVal 275
Db 783 AACATTAAAGTACACAG-----TCTGCACA 809
Qy 276 AlaThrLeuAsnIleGlyTyrPheGlyGlyIleGlyAlaArgLeuThrPhe 293
Db 810 CTAACATTATGATGTGTGCGCATTTCCGGTTAGAGATTGGAAGTATTTTCTTT 863

RESULT 18
US-09-660-587-43
; Sequence 43, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 43
; LENGTH: 828
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: nucleic acid sequence of E. canis p28-3
US-09-660-587-43

Alignment Scores:
Pred. No.: 4,6e-43 Length: 828
Score: 423.50 Matches: 109
Percent Similarity: 51.84% Conservative: 46
Best Local Similarity: 36.45% Mismatches: 115
Query Match: 28.31% Indels: 29
DB: Gaps: 9

US-10-062-624-40 (1-293) x US-09-660-587-43 (1-828)
Qy 1 MetAsnAsnIlyLeuPheHisIleLeuAsnThrValLeuValCysLeuLeuSerIleu 20
Db 1 ATGAACGTGTA--AAATTTCTTATTAACACTACATTGCTATCACTAACATCTTTTA 57
Qy 21 ProAsnIleSerSerSerIlyValAlaIleAsnAsnAlaIlyValIlyTyrGlyLeuTyr 40
Db 58 CCTGCATATCTTTTCTCCAAACCATATCATGAAAC--AATCTACAGGAAACTTTTAC 114
Qy 41 IleserGlyGlnTyrIlyProSerValSerValPheSerAsnPheSerValIlyGlnThr 60

```

```

Db 115 ATTATGGAAAATATGATGACCAAGTATTTCACTTTTGGAACTTTTCAGTAAAGAA 174
Qy 61 AsnValIleIthrIlyAsnLeuIleAlaLeuIlyIysAspValAspSerIleGluThrIys 80
Db 175 AAAACACACACAACTGAAATTTTGGATTA--GAACTACAG 216
Qy 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIlePro---TyrThrAla 99
Db 217 ACTGCTGTATCATCTCTGATTAAGACATGCACTTTTATATATCCAAATATTCATTT 276
Qy 100 ValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGlu 118
Db 277 AAATATGAAATATATCATCTTTTACGATTTTGCAGGGTAAATGGCTATTAATAGT--- 333
Qy 119 GlyThrArgValGluIleGluGlySerTyrGlyGluPheAspValIlyAsnProGlyGly 138
Db 334 AGTCCAAAGATAGATTGGAAGTATCAATGACAGACATTCGATGACAAATCCAGAGAT 393
Qy 139 TyrThrLeuSerAspAlaTyrArgTyrPheAlaIleuAlaArgIly-----Met 154
Db 394 AAGTTTAACTATGATGACATTAAGTATGTGCTTTATCAATGATTCAGTAAACAAATG 453
Qy 155 IlyGlyAsnSerPheThrProIyGlyIlyValSerAsnSerIlePheHisIthrValMet 174
Db 454 AAAAGTGTAAATTC-----GTTTTC 477
Qy 175 ArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysTyrAspPheSerIleu 194
Db 478 AAAATGAAAGATTAAGACATATCACTGATTAATGATTAATGATTAATTAATTAAC 537
Qy 195 AsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhe 214
Db 538 AAAAGATGCTTTTTCACCTTACATATGTCAGGACATGCTGACTTATTAATTCATG 597
Qy 215 PheAspValLeuHisIleIysPheAlaTyrGlnSerIlyLeuGlyIleAlaTyrSerIleu 234
Db 598 TTTGACGCTATTAACCAATTAAGCTGCTTATCAAGAAATTAAGTTTAATTCATTA 657
Qy 235 ProSerAsnIleSerIleuPheAlaSerIleuTyrTyrHisIlyValMetGlyAsnGlnPhe 254
Db 658 AGCCAGAGACTAATTTCTATGCGTGTGCACTTTCAAAAGTAAACAAACAGAGTTT 717
Qy 255 LysAsnLeuAsnValGlnHisIValAlaGluLeuAlaSerIleProIyalleIthrSerVal 274
Db 718 AGAGTTCTGTCTTATTAAGCTGAGAGACTCGT-----CCAGATATATCTATTTC 771
Qy 275 ValAlaThrLeuAsnIleGlyTyrPheGlyGlyIleGlyAlaArgLeuThrPhe 293
Db 772 ATAGTAAAGTTGAGATATATGTCATTTTGGTTAGAAATTTGGGTACAGGCTGATTT 828

RESULT 19
US-08-953-326-11
; Sequence 11, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganca, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Susan M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent In Ver. 2.0

```

SEQ ID NO 11  
 LENGTH: 830  
 TYPE: DNA  
 ORGANISM: Ehrlichia chaffeensis  
 US-08-953-326-11

## Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	7,086-43	830	107	57	99	38
Percent Similarity:	422.00					
Best Local Similarity:	54.498					
Query Match:	35.558					
DB:	28.214					
	4					13

US-10-062-624-40 (1-293) x US-08-953-326-11 (1-830)

```

QY      1 MetAenAenLysLeuLysPheThrIlelleAenThrValLeuValCysLeuLeu---Ser 19
        |||||
        3 ATGAATTACAAAAGTTTTC-----ATACAGAGCGATGATATCATTAATATCTTCT 56
        |||||
QY      20 LeuProAenLysSerSerSer-----LysAlaIleAenAenAenAlaLysLys 35
        |||||
        57 CTACCTGGAGATCATTTTCCAGCCAGAGAGTAGTGATTAACGGTAAT-----107
        |||||
QY      36 TyrTyrGlyLeuTyrIleSerGlyGlnTyrLysProSerValSerValPheSerAenPhe 55
        |||||
        108 -----TTCTACATCATGCGAATAATCATGCCAAGTCCTTCGATTTTGAGATATTC 158
        |||||
QY      56 SerValLysGlyLysThrAenValIleThrLysAenLeuIleAlaLeuLysLysAenValAen 75
        |||||
        159 TCTGCTAAGAGAAAGAAATACACAGTGGAGTGTGGAGCTGGAGCTGAGCAAAATGGGAGC 218
        |||||
QY      76 SerIleGlyThrLysThrAenAenAenValGlyIleSerAenProSerAen---PheThr 94
        |||||
        219 GGA-----AGCGAATATCCAACTCTCCCAACGATGATATCTACT 260
        |||||
QY      95 IlePro---TyrThrAlaValPheGlnAenAenSer---ValAenPheAenGlyThrIle 112
        |||||
        261 GGTCAAAATTAATCATTAATAATGAAACACCCGTTTATGATTTGCAGAGCTAT 320
        |||||
QY      113 GlyTyrThrPheAlaGlyLysThrArgValGluIleGlyLysTyrGluGluPheAen 132
        |||||
        321 GGTTACTCATG---GATGCTCAAGAAATAGAGCTTGAAGTATCTTATCAACATTTGAT 377
        |||||
QY      133 ValLysAenProGlyLysThrLysSerAenAenAlaTyrArgTyrPheAlaLeuAlaArg 152
        |||||
        378 GTRAAATAATCAAGATTAACAATTAAGAATGAAGCATAGATATGCTCTATCCCAT 437
        |||||
QY      153 GluMetLysGlyAenSerPheThrProLysGlyLysValSerAenSerIlePheHisThr 172
        |||||
        438 -----AACTCAGCAGACAGCATGAGTACGCAAGTAAATATTT-----GTC 479
        |||||
QY      173 ValMetArgAenAenGlyLysSerIleIleSerValIleValAenValCysTyrAenPhe 192
        |||||
        480 TTTCTAAATAATCAAGATTAATCTTGAATATCTTATGCTGAACCAATGATGACGTA 539
        |||||
QY      193 SerLeuAenAenLysSerIleSerProTyrIleCysGlyValAlaGlyValAenAlaIle 212
        |||||
        540 GTRAGCGAAGCATACCTTTTCTCTATATATAGCGCAGATACCGTACTGATTTAGTA 599
        |||||
QY      213 GluPheAenAenValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyr 232
        |||||
        600 TCCATGTTTGAAGCTCAAAATCTTAATTTCTTACCAAGAAAGTATGATTTAGCTAC 659
        |||||
QY      233 SerLeuProSerAenLysSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAen 252
        |||||
        660 TCTATTAAGCCCAAGACCTCTGTGTATTTATGTGGGCACTTTCATAGGTAATAGGAGAC 719
        |||||
QY      253 GluPheLysAenLeuAenVal-----GlnHisValAlaGluLeuAlaSer 267
        |||||
        720 GAATTTAGATATTTCTCTATTAATTAATCTGATGATCAACCTTGCAGAAAGAAAGAAC 779
        |||||
QY      268 IleProLysIleThrSerAlaValAlaThrLysAenIleGlyTyrPheGlyGlyIle 287
  
```

DB 780 TACCCT-----GCATAGTAATACTGATGATATCCACTTTGATAGTAATG 827

QY 288 Gly 288  
 DB 828 GGA 830

## RESULT 20

US-09-648-520E-48  
 Sequence 48, Application US/09648520E  
 Patent No. 6432649  
 GENERAL INFORMATION:  
 APPLICANT: Stich, Roger W.  
 TITLE OF INVENTION: Methods for Detecting Ehrlichia Canis and Ehrlichia Chaffeensis  
 TITLE OF INVENTION: Vertebrate and Invertebrate Hosts  
 FILE REFERENCE: 22727/04069  
 CURRENT FILING DATE: 2000-08-25  
 NUMBER OF SEQ ID NOS: 51  
 SOFTWARE: PatentIn version 3.1

SEQ ID NO 48  
 LENGTH: 849  
 TYPE: DNA  
 ORGANISM: ORF sequence, Ehrlichia chaffeensis p28  
 US-09-648-520E-48

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	6,216-42	849	103	47	121	33
Percent Similarity:	414.50					
Best Local Similarity:	49.34%					
Query Match:	27.71%					
DB:	4					10

US-10-062-624-40 (1-293) x US-09-648-520E-48 (1-849)

```

QY      1 MetAenAenLysLeuLysPheThrIlelleAenThrValLeuValCysLeuLeu---Ser 19
        |||||
        1 ATGAATTACAAAAGTTTTC-----ATACAGAGCTGATGATATCATTAATATCTTCT 54
        |||||
QY      20 LeuProAenLysSerSerSer-----LysAlaIleAenAenAenAlaLysLys 35
        |||||
        55 CTACCTGGAGATCATTTTTCGACCCACAGAGTGTGATTAACGGTAAT-----105
        |||||
QY      36 TyrTyrGlyLeuTyrIleSerGlyGlnTyrLysProSerValSerValPheSerAenPhe 55
        |||||
        106 -----TTCTAATATCAGTGAATAATATGCAAGAGCTTGCATTTTGGGCTRTTY 156
        |||||
QY      56 SerValLysGlyLysThrAenValIleThrLysAenLeuIleAlaLeuLysLysAenValAen 75
        |||||
        157 TCTGCTAAGAGAAAGAAATACACAGTGGAGTGTGGAGTATGATGACCAATTTGGAG 216
        |||||
QY      76 SerIleGlyThrLysThrAenAenAenValGlyIleSerAenProSerAenPheThrIle 95
        |||||
        217 GMAAGYCAATATACACACACACACACACACACACACACACACACACACACACACACAC 276
        |||||
QY      96 ProTyrThrAlaValPheGlnAenAenSer---ValAenPheAenGlyThrIleGlyTyr 114
        |||||
        277 AAATAT-----GMAAAATATATATATATATATATATATATATATATATATATATAT 324
        |||||
QY      115 ThrPheAlaGlyLysThrArgValGluIleGlyLysTyrGluGluPheAenValAen 134
        |||||
        325 TCAATG---GATGCGCAAGAAATAGAGCTTGAAGATCTTATGACCAATTTGATGATG 381
        |||||
QY      135 AenProGlyLysTyrThrLysSerAenAenAlaTyrArgTyrPheAlaLeuAlaArgGluMet 154
        |||||
        382 AATCAAGGTAACATATATATATATATATATATATATATATATATATATATATATATAT 441
        |||||
QY      155 LysGlyAenSerPheThrProLysGlyLysValSerAenSerIlePheHisThrValMet 174
        |||||
        442 WCABBARCARRCATGMSK-----AGTGCAARRATATATATATATATATATATATATAT 486
  
```

```

Oy 175 ArgAnaAaAGGLLeuSerIleIleSerValIleValAsnValAlcYstYrAspPheSerLeu 134
Db 487 AAAAAATGAAGGNTACTTGAACRTATCTTATGCTGAACGACGCTATGATGATARGY 546
Oy 195 AsnAenLeuSerIleSerProTYrIleCYsGlyAlaGlyValAspAlaIleGluPhe 214
Db 547 GAAGGATATCCCTTTTCTCTCTTAATATGAGTGAAGTATGGAACCTGATTTAGATCCATG 606
Oy 215 PheAspValLeuHisIleIleLysPheAlaTYrGlnSerLysLeuGlyIleAlaTYrSerLeu 234
Db 607 TTGGAAGYACAAAAAYCCTAAATTTCTTACCAAGAGAAAGTTAGTTAAGCTACTCATATA 666
Oy 235 ProSerAenIleSerLeuPheAlaSerLeuTYrTYrHisIleValMetGlyAsnGlnPhe 254
Db 667 AGCCAGAAACCTCTCTGTTTATGTYGGYGGACATTTTCATTAAGGTATATGGAACGAATY 726
Oy 255 LysAenLeuAsnVal-----GlnHisValAlaGluLeuAsnIleSerIlePro 269
Db 727 AGAGATATTCCTCCTCTTAATACTTAVGTGATCAASCTTGACAGAAAGAAATYACCT 786
Oy 270 LysIleIleHisSerAlaValAlaThrLeuAsnIleGlyTYrPheGlyGlyGluIleGlyAla 289
Db 787 -----GCATAGTAATAVACTRGAATGATGCACTTGGMATAGACCTTGAGAGA 834
Oy 290 ArgLeuThrPhe 293
Db 835 AGRTTTCCTTC 846

RESULT 21
US-08-733-230-3
Sequence 3, Application US/08733230
Patent No. 6025338
GENERAL INFORMATION:
APPLICANT: Barbet, Anthony F.
APPLICANT: Ganta, Roman Reddy
APPLICANT: McGulire, Travis C.
APPLICANT: Burridge, Michael J.
APPLICANT: Nyika, Aceme
APPLICANT: Rurangirwa, Fred R.
APPLICANT: Mahan, Suman M.
TITLE OF INVENTION: Nucleic Acid Vaccines Against
TITLE OF INVENTION: Rickettsial Diseases and Methods of Use
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSER: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/733,230
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF-167
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-375-5800
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 842 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

```

FEATURE	NAME/KEY:	CDS	LOCATION:	1..840
US-08-733-230-3				
Alignment Scores:				
Pred. No.:	2.54e-41	Length:	842	
Score:	409.50	Matches:	105	
Percent Similarity:	53.62%	Conservative:	58	
Best Local Similarity:	34.54%	Mismatch:	102	
Query Match:	27.37%	Indels:	39	
DB:	3	Gaps:	13	
US-10-062-624-40 (1-293) x US-08-733-230-3 (1-842)				
QY	1 MetAsnAsnIysLeuIysPhe--- <td>19</td> <td>     </td> <td>     </td>	19		
Db	1 ATGATATTACAAAAGAAAGTTTCATTACACAGCATGAT-----ATCATTAATATCTCTC	54		
QY	20 LeuProAsnIleSerSerSerIys-----AlaIleAsnAsnAlaIys	34		
Db	55 TTACCTGAGATCATCTTTCCACCCAAAGCAGCAGTAGTGCATTAACGGTAAT-----	108		
QY	35 LyetYrTyGlyLeuTyrlleSerGlyGlnTyrlsProSerValSerValPheSerAsn	54		
Db	109 -----TTTCAATCACTAGTGGAAATTCAGATGCCAAGCCTTCCCATTTTGGAGTA	156		
QY	55 PheSerValIleGlyIuThrAsnValIleThrlsAsnLeuIleAlaLeuIysAspVal	74		
Db	157 TTCTCTGCTACGAAAGAAATACACACAGTGGAGTCTTTGGCATGACACAAATTTGG	216		
QY	75 AspSerIleGlyIuThrIysThrAspAlaSerValGlyIleSerAsnProSerAsn--Phe	93		
Db	217 GACGGA-----AGCGCAATATCAACTCTCCCAACAGATGTATTC	258		
QY	94 ThrIlePro---TyThrlAlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThr	111		
Db	259 ACCTGCTCAAAATTATTCATTAAATATGAAACAAACCGTTTATGATTGGTCAGAGACT	318		
QY	112 IleGlyTyThrPheAlaGlyIuThrlsIysValGlnIleGlyIysTyrlsGlyIuPhe	131		
Db	319 ATTTGGTATCTCAATG---GATGCTCCAAAGATAGAGCTTGAAGTATCTTATGAACATTT	378		
QY	132 AspValIysAsnProGlyIuTyThrlsIysSerAspAlaTyrlsGlyPheAlaLeuAla	151		
Db	376 GATGTAAATAAATCAAGGTAAATTAATTAAGATAGACACATAGATATTTGGCTCTATCC	433		
QY	152 ArgGlyMetIysGlyAsnSerPheThrProIysGlyIuValSerAsnSerIlePheHis	171		
Db	436 CAT-----AACTACGACACAGACATGATGACAAATTAATTAATTTT-----	477		
QY	172 ThrValMetArgAsnAspGlyLeuSerIlelleSerValIleValAsnValCysTyrlAsp	191		
Db	478 GCTCTTTCTAAAAAATGAAAGATTACTTGACATATCATTTATGTCGAAGCAGCTTAAGAC	537		
QY	192 PheSerIleAsnAsnLeuSerIleSerProTyrlleCysGlyIuAlaGlyValAspAla	211		
Db	538 GTAGTAGGGGAGGACATACCTTTTCTCTTATATATCCGAGGTATCCGATCATGATTA	597		
QY	212 IleGlyPhePheAspValLeuHisIleIysPheAlaTyrlsIysIysLeuGlyIleAla	231		
Db	598 GTATTCATGTTTGAAGCTACAAATCTCTAAAAATTTCTTACCAAGAAAGTTAGGTTTAAGC	657		
QY	232 TyrlsSerLeuProSerSerAsnIleSerIuPheAlaSerIleuTyrlsIysValMetGly	251		
Db	658 TACTCTATTAAGCCCAAGACCTTCTGTGTATTGTGTGGGCACTTTCATTAAGGTAAATGGG	717		
QY	252 AsnGlnPheIysAsnLeuAsnVal-----GlnHisValAlaGlyLeuAla	266		
Db	718 AACGATTTTAGAGATATTTCTACTATTAATACCTGATCAACACTTGCAGGAAAGGA	777		
QY	267 SerIleProIysIleThrSerAlaValAlaThrlsLeuAsnIleGlyTyrlsPheGlyIu	286		

```

Db      77  AACTACCC-----GCAATAGTAATACGTGATGATGCACTTGGAATGAA  825
Qy      287  ILGLYALATG 290
          : : : : :
Db      826  ATGGAGGAAG 837

RESULT 22
US-08-953-326--3
; Sequence 3, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Gantle, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rutangira, Fred R.
; APPLICANT: Mahan, Sunan M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167CI
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 842
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(840)
US-08-953-326-3

Alignment Scores:
Pred. No.:      2.54e-41      Length:      842
Score:          409.50      Matches:    105
Percent Similarity: 53.62%      Conservative: 58
Best Local Similarity: 34.54%      Mismatches: 102
Query Match:    27.37%      Indels:    39
DB:              4          Gaps:      13
US-10-062-624-40 (1-293) x US-08-953-326-3 (1-842)

Qy      1  MetAsnAsnLysLeuLysPhe---ThrIleIleAsnThrValLeuValCysLeuLeuSer  19
          : : : : :
Db      1  ATGATTAATACAAAAAAGTTTCATATACAGCGATTGAT-----ATCATTAATATCTCTCTC  54

Qy      20  LeuProAsnIleSerSerSerLys-----AlaIleAsnAsnAlaLys  34
          : : : : :
Db      55  TTACTTCGAGATATCATTTTCCAGCCCAAGCAGATAGTGTATTTAACGGTAAT-----  108

Qy      35  LysTyrTyTgLYLeuTYrIleSerIgLInTYrLYsProSerValSerValPheSerAsn  54
          : : : : :
Db      109  -----TTCTACATCACTGAGAAATACGAATGCCAAGCGTTCCGATTTGGAGTA  156

Qy      55  PheSerValLysGluThrAsnValIleThrLYsAsnLeuIleAlaLeuLYsLYsAspVal  74
          : : : : :
Db      157  TTCTCTCTCTAAGGAAAGAAATACAAATACAGTTGAGTGTTTGGACTGAACCAAAATTGG  216

Qy      75  AspSerIleGluThrLYsThrAspAlaSerValGLYIleSerAsnProSerAsn---Phe  93
          : : : : :
Db      217  GACCGA-----ACCGCAATATCCCACTCTCTCCCAAGATGTATTC  255

Qy      94  ThrIlePro---TYrThrAlaValPheGlnAspAsnSer---ValAsnPheAsnGLYThr  111
          : : : : :
Db      259  ACTCTCTCAATATATTCATTTAATATATGAAAAACAACCGCTTTTATGAGTTTGCAGAGACT  318

Qy      112  ILeelyTYrThrPheAlaGluGLYThrAspValGluIleGluLYsSerTYrGluGluPhe  131

```

```

REFERENCE/DOCKET NUMBER: UF-167
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 849 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..846
US-08-733-230-5

Alignment Scores:
Pred. No.: 4,94e-27 Length: 849
Score: 294.00 Matches: 82
Percent Similarity: 45.42% Conservative: 57
Best Local Similarity: 26.80% Mismatches: 129
Query Match: 19.65% Indels: 38
DB: 3 Gaps: 9

US-10-062-624-40 (1-293) x US-08-733-230-5 (1-849)
QY 1 MetAanLnLYleuLpPheThr-----IlelleAnThrValleuValCysleu 17
DB 1 ATGAATTACAGAAATTGTTTACAGGGGCGCTGTACAGACCCAGCTGCGCTGCTCC 60
QY 18 LeuSerLeuProAenlleSerSerlyAlaIleAenAenAlaIlelySerlyTyx 37
DB 61 CTACTGTATTAGGGGCGCTAGTGCATCTCCATGAGCAGCAAGTGGCTTGAAGG 120
QY 38 Glyleu-----TyrIleSerGlylnTyxPProSerValSerValPhe 52
DB 121 GGAGTAATGGAGTAGCTTTTACGTCGGCTCAGCCAGCCAGCTTCTCTGTT 180
QY 53 SerAenPheSerVallyleGlnThraenValIleThrlyAenlleIleAlaIleuLylys 72
DB 181 ACCTGCTTCGACATGCTGAGTCAGCAAGACCTCATACGTAGAGGCTATGACAG 240
QY 73 AspAenAspSerlleGlnThrlySerThraAspSerValGlyIleSerAenProSerAen 92
DB 241 AGCATT-----GCCAGATTGAGTGAAGTGCACAGAAC 276
QY 93 PheThr---lleProTyThraValPheGlnAenAenSerVal---AenPheAenGly 110
DB 277 TTTTCCAATCTGGCTACACTTTTCCCTCTCTAAAACTTAATCAAGCTTTCGACG 336
QY 111 ThrIleGlyTyThrPheAlaGlnGlyThraGlnValGlnIleGlnGlySerTyxGlnGly 130
DB 337 GCTGTGGGATATCTCTGGGA---GAGCCAGAGTGAATGGAAGCGAGCTACAGAAAG 393
QY 131 PheAenVallyAenAenProGlyGlyTyThrIleSerAenAenAenAenAenAenAen 150
DB 394 TTTGCTACTTTGGCGGAGCGGAGTACGAAAAAGTGTGCGAATCTCTGCGACACTT 453
QY 151 AlaAenGlyl-----MetlyGlyAenSerPheThrProLyGlnLyValSerAen 167
DB 454 ACCCCGCGAGCTACATTACTGAGCAATTAATCTTC----- 489
QY 168 SerIlePheHleThrValMetAraGlnAenAenAenAenAenAenAenAenAenAen 187
DB 490 -----GTAGTCAAAATTGATGAAATCACAACACCTTCAGTCAATGTTAAT 534
QY 188 ValCysTyxAspPheSerSerlyAenAenAenAenAenAenAenAenAenAenAen 207
DB 535 GCGCTGTATGAGCTGTGACACACAGATTACTGTCTCCCTATGATATGTCGGGATA 594
QY 208 GlyValAenAenAenAenAenAenAenAenAenAenAenAenAenAenAenAen 227
DB 595 GCGCGAAGCTTTGATGATCTCTAAGCAAGTAAACCAAGAGCTGCGCTACAGGGGCAAG 654

```

```

QY 228 LeuGlyIleAlaTyxSerLeuProSerAenlleSerleuPheAlaSerleuTyxThis 247
DB 655 GTGGATTACTTACCACTTACCTCCGGAATATCTCTGGGAGCGGGGCTTCTACAC 714
QY 248 LysValMetGlyAenGlnPheLyAenAenAenAenAenAenAenAenAenAenAen 267
DB 715 GGGCTATTGATGAGTCTTACAGAGCATCTCCGACACAGAGTAAAGTCTCTGGA 774
QY 268 IleProLyIleThrSerAlaValAlaThreAenlleGlyTyxPheGlyGlyIle 287
DB 775 GAAGCAAAA-----GCCTCAGTCAMAGCGATATGCTGACTAGGCTTAACTT 825
QY 288 GlyAlaAenGlyleuThrPhe 293
DB 826 GGAGCAAGATTCTCTGTT 843

RESULT 24
US-08-953-326-5
Sequence 5, Application US/08953326
Patent No. 6251872
GENERAL INFORMATION:
APPLICANT: Bardet, Anthony F.
APPLICANT: Ganata, Roman R.
APPLICANT: McGuire, Travis C.
APPLICANT: Burridge, Michael J.
APPLICANT: Nyika, Aceme
APPLICANT: Rutangitwa, Fred R.
TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
FILE REFERENCE: UF-167C1
CURRENT APPLICATION NUMBER: US/08/953,326
CURRENT FILING DATE: 1997-10-17
EARLIER APPLICATION NUMBER: 08/953,326
EARLIER FILING DATE: 1997-10-17
EARLIER APPLICATION NUMBER: 08/733,230
EARLIER FILING DATE: 1996-10-17
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 5
LENGTH: 849
TYPE: DNA
ORGANISM: Anaplasma marginale
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(846)
US-08-953-326-5

Alignment Scores:
Pred. No.: 4,94e-27 Length: 849
Score: 294.00 Matches: 82
Percent Similarity: 45.42% Conservative: 57
Best Local Similarity: 26.80% Mismatches: 129
Query Match: 19.65% Indels: 38
DB: 4 Gaps: 9

US-10-062-624-40 (1-293) x US-08-953-326-5 (1-849)
QY 1 MetAanLnLYleuLpPheThr-----IlelleAnThrValleuValCysleu 17
DB 1 ATGAATTACAGAAATTGTTTACAGGGGCGCTGTACAGACCCAGCTGCGCTGCTCC 60
QY 18 LeuSerLeuProAenlleSerSerlyAlaIleAenAenAlaIlelySerlyTyx 37
DB 61 CTACTGTATTAGGGGCGCTAGTGCATCTCCATGAGCAGCAAGTGGCTTGAAGG 120
QY 38 Glyleu-----TyrIleSerGlylnTyxPProSerValSerValPhe 52
DB 121 GGAGTAATGGAGTAGCTTTTACGTCGGCTCAGCCAGCCAGCTTCTCTGTT 180
QY 53 SerAenPheSerVallyleGlnThraenValIleThrlyAenlleIleAlaIleuLylys 72

```

```

Db      181 ACCCTGTCGATCGTGTGATCAAGCAAGACCTCATAGCTATGAGCTATGACAAAG 240
QY      73 AspvAlaSerIleGluThrIleThrAspAlaSerValGlyIleSerAspProSerAsn 92
Db      241 AGCATT-----GCACGATTGATGTAGTGTGACCAAC 276
QY      93 PheThr---IleProTyThrAlaValPheGlnAspAsnSerVal---AsnPheAsnGly 110
Db      277 TTTCCAAATCGGCTACACCTTTGGCTTCTTAATAAATCATACAGCTTTTCAGACGC 336
QY      111 ThrIleGlyTyThrPheAlaGluGlyThrArgValGluIleGlyIleSerTyGluGlu 130
Db      337 GCTGTGGATATTCCTGGGA---GGAGCCAGATGTAATGTAAGCCGATCAAGAG 393
QY      131 PheAspValLysAsnProGlyGlyTyThrIleSerAspAlaTyArgTyPheAlaLeu 150
Db      394 TTTGCTACTTTGGCGGACGGAGTACGCAAAAAGGTGGCAATCTCTGCGACCTAT 453
QY      151 AlaArgGlu-----MetLysGlyAsnSerPheThrProLysGluLysValSerAsn 167
Db      454 ACCCGGACGCTAACCTTACTGAGCAATTTACTT----- 489
QY      168 SerIlePheHisThrValMetArgAsnAspGlyLysSerIleIleSerValIleValAsn 187
Db      490 -----GTAGTCAAAATTTGATGAATCATCAAAACCTCATGCTTAAAT 534
QY      188 ValCysTyArgPheSerIleLysAsnIleSerIleSerProTyIleCysGlyGlyAla 207
Db      535 GCGCTGTATGACGTCTGCAACAGATTATTCCTGTCTCCGATGATGATGCTCCGGATA 594
QY      208 GlyValAspAlaIleGluPhePheAspValIleuHisIleLysPheAlaTyGlnSerLys 227
Db      595 GCGGCAAGTTTGTGACATCTCTTAAGCAAGTAAACCAAGCTGCGCTACAGGGGCAAG 654
QY      228 LeuGlyIleAlaTySerIleuProSerAsnIleSerIleuPheAlaSerIleuTyThrHis 247
Db      655 GTTGGGATTAGCTACAGTTTACTCCGCAAAATATCTTGTGGCGAGTGGCTTACACAC 714
QY      248 LysValMetClyAsnGlnPheLysAsnLeuAsnValGlnHisValAlaGluLeuLysAsp 267
Db      715 GGGCTATTGATGATGCTTACAAAGACATTCGCGCACACACAGCTTAAGTTCTTGGA 774
QY      268 IleProLysIleThrSerAlaValAlaThrLeuAsnIleGlyTyArgGlyGlyGluLe 287
Db      775 GAAGCAAA-----GCTTCAGTCAAGGCAATTTGCTGACGACTTAACCTT 825
QY      288 GlyAlaArgLeuThrPhe 293
Db      826 GGAGCAAGATTCCTGTTTC 843

RESULT 25
US-09-288-339-3
/ Sequence 3, Application US/09288339
/ Patent No. 6436399
/ GENERAL INFORMATION:
/ APPLICANT: Rikihisa, Yasuko
/ APPLICANT: Zhi, Ning
/ APPLICANT: Chashi, No. 643639910
/ TITLE OF INVENTION: Nucleic Acid Encoding the Major Outer Membrane Protein of
/ TITLE OF INVENTION: the Causative Agent of Human Granulocytic Ehrlichiosis
/ FILE REFERENCE:
/ CURRENT APPLICATION NUMBER: US/09/288,339
/ NUMBER OF SEQ ID NOS: 44
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 3
/ LENGTH: 1278
/ TYPE: DNA
/ ORGANISM: P44-2
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(1278)

```

```

US-09-288-339-3
Alignment Scores:
Pred. No.: 5,1e-20 Length: 1278
Score: 239.50 Matches: 92
Percent Similarity: 34.13% Conservative: 50
Best Local Similarity: 22.12% Mismatches: 113
Query Match: 16.01% Indels: 161
DB: 4 Gaps: 14

US-10-062-624-40 (1-293) x US-09-288-339-3 (1-1278)
QY      28 AlaIleAsnAsnAlaValLysTyThrIleuTyIleSerGlyIleTyLysPro 47
Db      61 GCTTTGAGATGGTGGCGGATATTC-----TATGTGGTTGATGATGATGCA 114
QY      48 SerValSerValPheSerAsnPheSerValLysGluThrAsnValIleThrLysAsnLeu 67
Db      115 GCGTTTACCAAGATATGAGATTTTATGATTAAGGAGATTAACGAGACAAAGCAGTA 174
QY      68 IleAlaLeuLysLysAspValAspSerIleGluThrIleThrAspAlaSerValGlyIle 87
Db      175 TATCCATATCTTAAGAGATGCAAGAGGTGAACCTTGAG----- 213
QY      88 SerAsnProSerAsnPheThrIleProTyThrAlaVal---PheGlnAspAsnSer--- 105
Db      214 TCGCACAAGTTTGAATGCAACACACCTGATCTCGATGGATTGGCTTAAGCAACATGCTT 273
QY      106 ValAsnPheAsnGlyThrIleGlyTyThrPheAlaGluGlyThrArgValGluIleGlu 125
Db      274 GTAGCTATGAAAGGACAGTGTGTTATGTTAGTGT---GGTGCAGGGGTGAGCTAG 330
QY      126 GlySerTyGluGluPheAspValLys-----AsnProGlyGlyTyThrLysSerAsp 143
Db      331 ATTGTTTACGACCGCTTCAACACCAAGGTTATTGAGATAGTGTGATGAAGAAATGAA 390
QY      144 AlaTyArgTyThrPheAlaLeuAlaArgGluMet----- 154
Db      391 GGTGATACAGATATCTCTAGTAAAGATGTTAGCTTATGATGTTTACTGACAGACT 450
QY      154 ----- 154
Db      451 GATAGCTTACCGCTCTCTTCCCAAGACCTCGGTAAAGATTCGTTCACTTGGCAAT 510
QY      154 ----- 154
Db      511 GCTGCAAAATTTCTAGCTCTGCATGATGAGGAAGTTTGTACTGTGATGCAATGCTGAC 570
QY      154 ----- 154
Db      571 CTAGCGCTGTGACGAATGCGGGGMAAAAGTTGTTGTGAACCGGAAGCAGCGGAGT 630
QY      154 ----- 154
Db      631 ACTGATGGGATACGTCAACAGTGTAGTGTAGGCACTAGTGTGTTACACAGAAAT 690
QY      154 ----- 154
Db      691 CCGAAGTTATTTAGTACTTTTGTGACACTGTGAAGATTGCTGAGAGTAAAACTGGCCG 750
QY      155 ----- 158
Db      751 ACGGCAAGGCAAAATGCAACATCACTGAAGACGGGTGATCTATAGTAAACCCAAA 810
QY      159 ----- 167
Db      811 GCGGTGTACAGACCTAGTACAGAGACTTAACCTTGAAGAAAAAACATAGTACAGAG 870
QY      168 SerIlePheHisThrValMetArgAsnAsp-----GlyLysSerIleIle 182
Db      871 TTACTAGCTTAAGATTTGAAGGGGTGAAGTTGTTGATCAAGGCGGTTTCTTACT 930
QY      183 SerValIleValAsnValCysTyArgPheSerIleuAsnLeuSerIleSerProTy 202

```





Db 1153 TGTGTGCTTGGCGGTAAGTCTGCGGCGGTGATGAGC---CATATCACTCTTAAG 1209  
 Qy 222 PheAlaTYrGlnSerLeuGlyIleAlaTYrSerLeuProSerAsnIleSerLeuPhe 241  
 Db 1210 CTGTGTTAGATTAAGGCTGGGTGAGTTATGAGCTCTCTCTTAATCTCCGCTTTT 1269  
 Qy 242 AlaSerLeuTYrTYrHisValMetGlyAsn---GlnPheValAsnLeuAsnValGln 260  
 Db 1270 GCGGCTGATCTTACATGCGCTGTGTGGAGATGCGCTTTATGATGATCTGCCGCTCA 1329  
 Qy 261 HisValAlaGluLeuAlaSerIleProIysIleThr-----SerAlaValAlaThrLeu 278  
 Db 1330 CGCTTGTATGATGATAGTACTAGTCCGCGCGGTCTACTAAGATGACTGTTAGCTTAATCTTC 1389  
 Qy 279 AsnIleGlyTYrPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293  
 Db 1390 TCCATGCTTATGTCGGTGGGAATTTGTGTAGGTTTCTTTT 1434  
 RESULT 27  
 US-09-066-047-12/C  
 Sequence 12, Application US/09066047A  
 Patent No. 6306394  
 GENERAL INFORMATION:  
 APPLICANT: MURPHY, Cheryl  
 STOREY, James  
 BELTZ, Gerald A.  
 COUGHLIN, Richard T.  
 TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND METHODS OF  
 USE OF GRANULOCYTIC ERHLICHIA  
 NUMBER OF SEQUENCES: 41  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: HALE AND DORR LLP  
 STREET: 60 State Street  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: United States  
 ZIP: 02109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/066,047A  
 FILING DATE: 24-Apr-1998  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/044,869  
 FILING DATE: 25-APR-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Superko, Colleen  
 REGISTRATION NUMBER: 39,850  
 REFERENCE/DOCKET NUMBER: 106,941.156  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 526-6000  
 TELEFAX: (617) 526-5000  
 INFORMATION FOR SEQ ID NO: 12:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1843 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
 US-09-066-047-12  
 Alignment Scores:  
 Pred. No.: 2,43e-19 Length: 1843  
 Score: 236.00 Matches: 91  
 Percent Similarity: 33.98% Conservative: 50  
 Best Local Similarity: 21.93% Mismatches: 114

Query Match: 15.78% Indels: 160  
 DB: 4 Gaps: 14  
 US-10-062-624-40 (1-293) x US-09-066-047-12 (1-1843)  
 Qy 28 AlaIleAsnAsnAsnAlaLysLysTYrTYrGlyLeuTYrIleSerGlyGlnTYrLysPro 47  
 Db 1621 GCTTTGAGACTGGTGTGGCGGATATTTCT---TATGTGGCTTGATTAAGCTCA 1568  
 Qy 48 SerValSerValPheSerAsnPheSerValLysGluThrAsnValIleThrLysAsnLeu 67  
 Db 1567 GCGTTTGACAAGATTAAGATTTAGTATAGGAGATTAACGAGACTTAAGCACTGA 1508  
 Qy 68 IleAlaLeuLysLysAspValAspSerIleGluThrIleThrAspAlaSerValGlyIle 87  
 Db 1507 TATCATACTTAAAGATGGAAGAGCTAAAGCTAAG----- 1469  
 Qy 88 SerAsnProSerAsnPheThrIleProTYrThrAlaVal---PheGlnAspAsnSer-- 105  
 Db 1468 TCACACAAGTTTGACTGGAACACTCGATCTCGATTTGGCTTTAAGACACATGCTT 1409  
 Qy 106 ValAsnPheAsnGlyThrIleGlyTYrThrPheAlaGluGlyThrArgValGluIleGlu 125  
 Db 1408 GTACCTATGGAAGGTAGGTGGTTATGATGATTTGCT---GGTCCAGGGTTGACCTTAG 1352  
 Qy 126 GlySerTYrGluGluPheAspValLys-----AsnProGlyGlyTYrThrLeuSerAsp 143  
 Db 1351 ATTGTTACGAGCCCTTCAAGACCAAGGTTATAGATATGATGTGTAGTAAGAAAGATGA 1292  
 Qy 144 AlaTYrArgTYrPheAlaLeuAlaArgGluMet----- 154  
 Db 1291 GGTATACAGATATATCTACTAGCTAAGAGATTAGCTTATGATGTTACTGCGCAGACT 1232  
 Qy 154 ----- 154  
 Db 1231 GATAACTGTCTGCTCTCTTCCCAAGACCTGTGTAAGATATTGTTCACTTGTCTAAG 1172  
 Qy 154 ----- 154  
 Db 1171 GCGGTGGGTTTCTCATCCCGGTATTGATAGAAGTTGTGATGGGGTCATGACACG 1112  
 Qy 154 ----- 154  
 Db 1111 GAAAAAAGAGTGAAGATATATGCTCGCTGCGCAGACTATACGATGTGGCGCTACAG 1052  
 Qy 154 ----- 154  
 Db 1051 ACGAATTAAGACGCTCAGTGTATGTGGAAACCGCAAGCCGCGCAGAGAGATTG 992  
 Qy 154 ----- 154  
 Db 991 GCGTTGACTGATTTGTTAACAAACAAGTTGGAGAGTAAAGAAATTGGCCAAACGGGG 932  
 Qy 155 ----- 158  
 Db 921 TACGTTATGATGCGACAAAGTTATGCTCGCGCATAGAAATGTTAAAGCCGAGAGCC 872  
 Qy 159 -----PheThrProIysGlyLys-----ValSerAsnSer 168  
 Db 871 GTAGCTAAGACCTAGTACAGAGACTACCCCTTAAGAAAAACCATAGTAGCAGGGTTA 812  
 Qy 169 IlePheHisThrValMetArgAsnAsp-----GlyLeuSerIleIleSer 183  
 Db 811 CTAGCTAAGACTATTTGAAGGGGGTGAAGCTTTTGAATGATCAGAGCGGCTTTCTTACTTCC 752  
 Qy 184 ValIleValAsnValCysTYrAspPheSerLeuAsnAsnLeuSerIleSerProTYrIle 203  
 Db 751 GTAATGTCATATGCTGTATGATCTTCTTATGTAAGGTTAGAGTGTCTTATCTATCT 692  
 Qy 204 CysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIle-----Lys 221  
 Db 691 TGTGTGTCTTGGCGGTAAGTCTGTGGCGGTGATGAGC---CATATCACTCTTAAG 635

```

Oy 222 PheAlaTYRGIaSerIySeuIeugIyIleAlaTYRSeuIeProSeuAnIleSeuIePhe 241
Db 634 CTGCTTATAGATTAAAGGCTGGCTGGATGATTACAGCTCTCTCTGAATCTCCGCTTT 575
Oy 242 AlaSerLeuTYRTyrHisIyValMetGIaen---GlnPheIySaenIeuaenValGln 260
Db 574 GGGGGTGATCTTACCATCGCGCTGTGGAGATGGCGTTATGATGATCTGCCGGCTCA 515
Oy 261 HisValAlaGluLeuAlaSerIlePheIyPheIeThr-----SerAlaValAlaThrLeu 278
Db 514 CGCTCTTAGAAGATACACTAGTCCGGCGGCTCGTACCAAGATCTGCTATGTCTTAACCTC 455
Oy 279 AnIleGIaTYRPhelGIyGluIleGIaIyAlaArgLeuThrPhe 293
Db 454 TCcATGCGCTTATGTCGCTGGGGAATTGGTGTTAGCTTTGCTTT 410

RESULT 28
US-09-066-046-28
/ Sequence 28, Application US/09066046A
/ Patent No. 6204252
GENERAL INFORMATION:
APPLICANT: MURPHY, Cheryl
STOREY, James
BELTZ, Gerald A.
COUGHLIN, Richard T.
TITLE OF INVENTION: CHARACTERIZATION OF GRANULOCYTIC
EHRUCHIA AND METHODS OF USE
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: HALE AND DORR LLP
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: United States
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/066,046A
FILING DATE: 24-Apr-1998
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Superko, Colleen
REGISTRATION NUMBER: 39,850
REFERENCE/DOCKET NUMBER: 106,941.155
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 526-6000
TELEFAX: (617) 526-5000
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 3435 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-09-066-046-28

Alignment Scores:
Pred. No.: 1,24e-16 Length: 3435
Score: 217.50 Matches: 84
Percent Similarity: 33.69% Conservativity: 43
Best Local Similarity: 22.28% Mismatches: 101
Query Match: 14.54% Indels: 149
DB: 4 Gaps: 12

```

QY	57	VallysgsluThrAsnValIlePhrIlysAsnLeuIleAlaLeuIleuLysuLysAspValAspSer	76
DB	179	ATAAAGGAGAGTAAACGAGAGACTAAGGACGATATATCCACTTAAAGATGAAGAGT	238
QY	77	IleGluThrLysThrAspAlaSerValGlyIleSerAsnProSerAsnPherThrIlePro	96
DB	239	GTAAAGCTAAGG-----TACACAAAGTTTACCTGGAAACACTCT	277
QY	97	TyrThrAlaVal--PheGlnAspAsnSer--ValAsnPheAsnGlyThrIleGlyTyr	114
DB	278	GATCCTCGAATTGGGTTTAAAGACAACATGCTTGATGCTATGAAGAGCAGTGGTTAT	337
QY	115	ThrPheAlaGluGlyThrArgValGluIleGlyLysSerTyrGluGluLysPheAspValLys	134
DB	338	GGTATTTGT--GGTCCAGGGTTAGCTTGAGATTGGTTACGAGCGCTTCAAGACCAG	394
QY	135	-----AsnProGlyGlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArg	152
DB	395	GGTATTAAAGATAGTACGTATGTAAGAAAGATGAAGCAGATACATATATCTACTAGTTAAG	454
QY	153	GluMet	154
DB	455	GAGTTAGCTTATGATGTTGTTACTGACAGACTGATTAACCTTGCCGCTGCTTGGCAAA	514
QY	154	-----	154
DB	515	ACCTCGGGAAAGACATCGTTCAGTTGGCAATGCTGTGMAAAATTTCTTACCTTAAATT	574
QY	155	-----LysGlyAsnSerPhe	159
DB	575	GATGAGCAGGTTTGTATATTAATAATCATACAGTGTGAATACGGGAAAGGACACCTTT	634
QY	160	ThrProLysGluLys	164
DB	635	AATCCAGATCCCAAGACAACCGAAGATATATACAGCGCAGTCAGTGGTTGAACACGAG	694
QY	164	-----	164
DB	695	GGAAAGAAATTAAGTTAGCGATTTTTCGAGGTGTAGTTTGAAGATATTAAGAAATTGG	754
QY	165	-----ValSerAsnSerIle	169
DB	755	CCTATGTCGACAGCTGGGAGAGCAGTGTGCTCCTGTGTGGGTGCATTTAATGTAAAT	814
QY	170	PheIsthrValMetArgAsn-----	176
DB	815	GCCAAAGCTAATGGCTAAGAGACTTAATGATCTTAATCCAGACGAAAAACCATAGTAGCA	874
QY	177	-----AspGlyLeuSerIle	181
DB	875	GGGTTACTAGCTAAACATTGAAAGGTGAGTGGATTGTTGAATTAAGCGCGGTTCTTCT	934
QY	182	IleSerValIleValAsnValCysThrAspPheSerLeuAsnAsnLeuSerIleSerPro	201
DB	935	ACTTCTGAAATGGTCAATGCTTGTTATATCTTCTTAAAGTGAAGCTTACGCGTGTCTCT	994
QY	202	TyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuIstIle---	220
DB	995	TACGTTGTTCGCGCTTCGAGAGTAACTTCGTGGCGCTGTTGATGG---CATATCACT	105
QY	221	---LysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSer	239
DB	1052	CCTAAGCTGCTGTATAGATTAAAGCTGGGTGAAGTTACACTCTCTCCCTAAATCTCC	111
QY	240	LeuPheAlaSerLeuTyrThrHisLysValMetGlyAsn---GlnPheLysAsnLeuAsn	258
DB	1112	GCTTTGTGGGGGATTTATATCATGCGCGTGGGAGATGCTGCTATCATATATCTTCCA	117
QY	259	ValGlnHisValAlaGluLeuAlaSerIleProLysIleThr-----SerAlaValAla	276
DB	1172	GCTCAAGCTCTTGATGATGATATAGTCCGGCGGGTGCATCAAGATATCTGATGTCT	123
QY	277	ThrLeuAsnIleGlyTyrPheGlyGlyGluIleGlyAlaArgLeuThrPhe	293

DB 1232 AACTCTCATGCGTTATGTCGTCGGGAAATTTGGTGTAGTTGCTTT 1282

RESULT 29

US-09-066-047-14

Sequence 14, Application US/0906047A

Patent No. 6306394

GENERAL INFORMATION:

APPLICANT: MURPHY, Cheryl

STOREY, James

BEITZ, Gerald A.

COUGHLIN, Richard T.

TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND METHODS OF USE OF GRANULOCYTIC ERHLICHA

NUMBER OF SEQUENCES: 41

CORRESPONDENCE ADDRESS:

ADDRESSEE: HALE AND DORR LLP

STREET: 60 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: United States

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/066,047A

FILING DATE: 24-Apr-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/044,869

FILING DATE: 25-Apr-1997

ATTORNEY/AGENT INFORMATION:

NAME: Superko, Colleen

REGISTRATION NUMBER: 39,850

REFERENCE/DOCKET NUMBER: 106,941.156

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 526-6000

TELEFAX: (617) 526-5000

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 3435 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

SEQUENCE DESCRIPTION: SEQ ID NO: 14:

US-09-066-047-14

Alignment Scores:

Pred. No.: 1,24e-16 Length: 3435

Score: 217.50 Matches: 84

Percent Similarity: 33.69% Conservative: 43

Best Local Similarity: 22.28% Mismatches: 101

Query Match: 14,548 Indels: 149

DB: 4 Gaps: 12

US-10-062-624-40 (1-293) x US-09-066-047-14 (1-3435)

QY 57 VallyAGLThrsanAlleThryAsnleuIleAlaLeuylsAspValAspSer 76

DB 179 ATAGGAGAGTAACGAGAGTAAGGAGTATTCATCTTAAGAGATGGAAGAGT 238

QY 77 lIeGlThryThrsAspAlaservalGlylIeserAsnProserAsnPhetThrllePro 96

DB 239 GTAAAGTACGAG 277

QY 97 TyrlleThralValAl--PheGlnAspAsnSer--ValAsnPhetAsnGlyThrlleGlyTy 114

DB 278 GATCTCGAGATTGGGTTTAAAGACACATGCTTGTAGCTATGAGACGAGTGGTTAT 337

QY 115 ThrPheAlaGluGlyThrsArValGluIleGluGlySerTyrgluGluPheAspValys 134

DB 338 GGTATTGGT---GGTCCAGAGGTTGAGCTTGAAGATTGGTTACGAGCGTTCAAGACCAAG 394

QY 135 -----AsnProGlyGlyTyThrsLeuSerAspAlaTyrglyTyPheAlaLeuAlaArg 152

DB 395 GGTATTAGATAGTAGGTAGTAGAGAGAGATGAGACAGATACAGATATATCTAGCTAAG 454

QY 153 GluMet----- 154

DB 455 GAGTTAGCTTATGATGTTGTTACTGACACAGATGATTAACCTTGCCGCTCTTGCCAAA 514

QY 154 ----- 154

DB 515 AACTCGGGAGAGACATGCTTCACTTGGCCAAATGCTGAAATTTCTTACCTTAAT 574

QY 155 -----LygIlyAsnSerPhe 159

DB 575 GATGACAGAGTTTGTAAATCAATACAGTGTGAATACGGGGAAGGACAAACCTTT 634

QY 160 ThrProlysgluTy----- 164

DB 635 AATCCAGATCCCAAGACCAACGAAATTAACAGCGACGTGAGTGAACACGAAG 694

QY 164 ----- 164

DB 695 GGAACGAATAGTTTAGCGATTTTGTGAGAGTGTAGTTGAAAGATTAAGATTTGG 754

QY 165 -----ValSerAsnSerIle 169

DB 755 CCTACTGTGACGCTGGAGAGACAGTGTGCTGCTGTGGTGATCTTAATAGTAT 814

QY 170 PheHisThrValMetArgAsn----- 176

DB 815 GCCAAGCTATGCTAGAGACCTAGTAGATCTTAATGACAGCAAAACCAATAGTACA 874

QY 177 -----AspGlyLeuSerIle 181

DB 875 GGGTTACTAGCTAAACATTTGAAAGTGTGAGTGTGTGAGATTAGAGCGGCTTCTCT 934

QY 182 lIeserValIleValAsnValCysTyrsPheSerLeuAsnAsnLeuSerIleserPro 201

DB 935 ACTTCTGATAGTCATAGCTTGTATGATGCTTCTTACTGAAAGCTGACGCTTGTCT 994

QY 202 TyrlleGlyGlyValaGlyValaAspAlaIleGluPhePheAspValLeuHisIle--- 220

DB 995 TACGCTTGTGTGCTTGTGAGTAACTTCGTGGCGCTTGTGTGATGGG---CATATCACT 1051

QY 221 ---LysPheAlaTyrglnSerTyserLeuGlyIleAlaTyserLeuProSerAsnIleser 239

DB 1052 CCTAAGCTTGTCTTATAGTTAAAGCTGTGGTTGAGTTATCACTCTCTCCGAATCTCC 1111

QY 240 LeuPheAlaSerLeuTyThrsIleValMetGlyAsn---GlnPheLyAsnLeuAsn 258

DB 1112 GCTTTGCTGGGAGATCTTATCATGCGCTGTGGAGATGTGTCTATGATGATCTTCCA 1171

QY 259 ValGlnHisValaIleGluLeuAlaSerIleProlyserIleThr-----SerAlaVala 276

DB 1172 GCTCAAGCTTGTAGAGATAGTACGTGCGCGGCTGTACTACAGATATGCTTATGCT 1231

QY 277 ThrLeuAsnIleGlyTyPheGlyGlyGluIleGlyAlaTyglLeuThrsPhe 293

DB 1232 AACTCTCATGCGTTATGTCGTCGGGAAATTTGGTGTAGTTGCTTT 1282

RESULT 30

US-09-288-339-1

Sequence 1, Application US/09288339

Patent No. 6436399

GENERAL INFORMATION:

APPLICANT: Rikinis, Yasuko

APPLICANT: Zhi, Ning

```

; APPLICANT: Ohashi, No. 64363910
; TITLE OF INVENTION: Nucleic Acid Encoding the Major Outer Membrane Protein of
; TITLE OF INVENTION: the Causative Agent of Human Granulocytic Ehrlichiosis
; TITLE OF INVENTION: and Peptides Encoded Thereby
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/288,339
; CURRENT FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1242
; TYPE: DNA
; ORGANISM: P44
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1242)
US-09-288-339-1

Alignment Scores:
Pred. No.: 7,63e-15 Length: 1242
Score: 197.50 Matches: 78
Percent Similarity: 32.08% Conservative: 41
Best Local Similarity: 21.02% Mismatches: 95
Query Match: 13.20% Indels: 157
DB: 4 Gaps: 12

US-10-062-624-40 (1-293) x US-09-288-339-1 (1-1242)
QY 28 AAlaIleAsnAAsnAlaLylsYrTyrgIlyLeuTyrlleSerGIyGlnTyrlsAspPro 47
DB 127 GCTTGGAGACTGCTGTCGCGGATATTTC-----TATGTTGGCTTGGATTACAGTCCA 180
QY 48 SerValSerValPheSerAsnPheserVallysgIuThrsanValIleThrlsAsnleu 67
DB 181 GCGTTTACAGATACATAGAGATTATTAAGGAGACGAGACGAGACTTAAGGACAGTA 240
QY 68 IleAlaLeuLylsAspValAspSerIleGluThrlsThrsAlaSerValGlyle 87
DB 241 TATCATCATCTTAAGAGATGAAGAGTGAAGCTAGAG----- 279
QY 88 SerAsnProSerAsnPheserThrlleProTyrlsAlaVal---PheGlnAspAsnSer--- 105
DB 280 TCACACAAAGTTTACCTGGAACACACACCTGATCCTCGGATTGGGTTTAAGACAACTGCTT 339
QY 106 ValAsnPhesAsnGlyThrlleGlyTyrlsPheAlaGluGlyThrsAlaGluIleGlu 125
DB 340 GTAAGTATGAAGAGGAGGAGTGGGTAACCTATTGGT---GGGCCAGGAGTTGAGCTTGA 396
QY 126 GlySerTyrgIuGluPheAspValIys-----AsnProGlyGlyTyrlsThrlsSerAsp 143
DB 397 ATTGTTTACGAGCCGTTTCAAGACTAAGGATTAAGATAGTGGTATGAAGAAAGATGAA 456
QY 144 AlaTyrlsArgTyrlsPheAlaLeuAlaArgGluMet----- 154
DB 457 GCTGATACAGTATATCTACTAGCTAAGAGATTAGATTGATTGTTACTGGGCACT 516
QY 154 ----- 154
DB 517 GATAAAGCTTGTCTGCTCTTCCCAAACTTCGCGTAAGACATGTTCACTTGGCTAAG 576
QY 154 ----- 154
DB 577 GCGGTGAGATCTCTTACCCTAGTATCATGAGGAAGGTTTGTAGTGAAGAAAGATGCGCG 636
QY 154 ----- 154
DB 637 CTTCAGCAAAACACGAAACGCGAGAAAGTAACGCGTTAGACCTGCGAAGCGCGAACA 696
QY 154 ----- 154
DB 697 GACGGAGACAGCTGCGCACTGATGAGTTGTTAGTAATGTTAGTGGGAGGCTGCTCATAAA 756
QY 154 ----- 154
```

```

DB 757 TATTAGTAGTATTGTAAAGCTTACGCGAGTGTGAAGGAAAAAACTGGCCTACTGCT 816
QY 155 -----LysGlyAsnSer----- 158
DB 817 AGATGAGTAAACAAACAGCAATAGCATTTGTAGTGGAGCACTTAATAGCAAAACGCG 876
QY 159 -----PheThrProGlyIlys-----ValSerAsnSer 168
DB 877 ATGGCCAAAGATCTTAAGTAAGAACTTAACACAGAGAGAAACCATATGTTGCCGGTTA 936
QY 169 IlePheHisThrValMetArgAsnAsp-----GlyLeuSerIleIleSer 183
DB 937 CTAGCTAAACTATTTGAAGAGGGGAGGCTTGTAGAAATTAAGGGGCTTCTTCACTTCT 996
QY 184 ValIleValAsnValCysTyrlsAspPheSerLeuAsnLeuSerIleSerProTyrlle 203
DB 997 GTAATGTCAAATGCTTGTATGATCTTCTTATGAGAGGTTTAAAGGTTTGGTTCCTATGCT 1056
QY 204 CysGlyGlyAlaGlyValAlaPalaIleGluPheAspValLeuHisIle-----Lys 221
DB 1057 TGTGTTGTTTAAAGTGGTAACTCTGTAGTGTGTTGTTGAGG---CATATCACTTCTAAG 1113
QY 222 PheAlaTyrgInserTyrlsLeuGlyIleAlaTyrlsSerLeuProSerAsnIleSerIlePhe 241
DB 1114 CTTCGTTATAGCTTAAGAGCTGCGCTTGAATATCAGCTCTCTCGTGAATCTCCGCTTT 1173
QY 242 AlaSerLeuTyrlsThrlsLysValMetGlyAsn 252
DB 1174 GCGGGTGGATTCTTACCATCGTGTGTGGAGAT 1206

RESULT 31
US-09-288-339-5
; Sequence 5, Application US/09288339
; Patent No. 6436399
; GENERAL INFORMATION:
; APPLICANT: Rikhisas, Yasuko
; APPLICANT: Zhi, Ming
; APPLICANT: Ohashi, No. 64363910
; TITLE OF INVENTION: Nucleic Acid Encoding the Major Outer Membrane Protein of
; TITLE OF INVENTION: the Causative Agent of Human Granulocytic Ehrlichiosis
; TITLE OF INVENTION: and Peptides Encoded Thereby
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/288,339
; CURRENT FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1176
; TYPE: DNA
; ORGANISM: P44-12
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1176)
US-09-288-339-5

Alignment Scores:
Pred. No.: 9.32e-15 Length: 1176
Score: 196.50 Matches: 78
Percent Similarity: 32.35% Conservative: 42
Best Local Similarity: 21.02% Mismatches: 94
Query Match: 13.14% Indels: 157
DB: 4 Gaps: 12

US-10-062-624-40 (1-293) x US-09-288-339-5 (1-1176)
QY 28 AAlaIleAsnAAsnAlaLylsYrTyrgIlyLeuTyrlleSerGIyGlnTyrlsAspPro 47
DB 61 GCTTGGAGACTGCTGTCGCGGATATTTC-----TATGTTGGCTTGGATTACAGTCCA 114
QY 48 SerValSerValPheSerAsnPheserVallysgIuThrsanValIleThrlsAsnleu 67
DB 115 GCGTTTACAGATACATAGATTATTAAGGAGATGAAGAGAGACTTAAGGACAGTA 174
```

```

QY 68 IleAlaLeuLYeAspValAspSerIleGluThrLYeThrAspAlaSerValGlyIle 87
DB 175 TATCCACTACTTAAAGATGGAAGAGTGAAGCTAGAG----- 213
QY 88 SerAsnProSerAsnPhenThrIleProTYrThrAlaVal---PheGlnAspAsnSer--- 105
DB 214 TCACACAAGATTGACTGGAAACACTCCTGATCTCGGATGGGTTTAAGGACAAACATGCTT 273
QY 106 ValAsnPhenAsnGlyThrIleGlyTYrThrPheAlaGluGlyTYrThrArgValGluIleGlu 125
DB 274 GTAGCTATGAGGAGGAGGCTTGATGATGATGCT---GGTCCAGGAGGCTTGAAG 330
QY 126 GlySerTYrGluGluPheAspValLYs-----AsnProGlyGlyTYrThrLeuSerAsp 143
DB 331 ATTGCTTACGAGCGCTTCAAGACCAAGGCTATTAGAGATAGCTGATAGTAAGAAATGA 390
QY 144 AlaTYrArgTYrPheAlaLeuAlaArgGluMet----- 154
DB 391 GCTGATACAGTATATCTACTACTAGAGATTAGCTTATGATGATGTTTACTGGCAGACT 450
QY 154 ----- 154
DB 451 GATAACCTTCTGCTGCTCTTCCAAAGACCTGTGTAAGATATTGTCAGTTGCTAAG 510
QY 154 ----- 154
DB 511 GCGGTTGGGTTTCTCATCCCGTATGATTAAGAAAGTTTGATGGGGGTATGCACCG 570
QY 154 ----- 154
DB 571 GGAAGAAAGATGAGATATAGCTCGCTGCGCGCACTATACGATGTTGGCGCTCACAG 630
QY 154 ----- 154
DB 631 ACGAATTAGACGCGCTCAGTGTATGTATGGAACCGCAAGCCGCGCAAGAGAGATTG 690
QY 154 ----- 154
DB 691 GGCTTGACTGAGTTTGTTAACAAGTTTGAAGAGTGAAGATTGGCCACCGGGG 750
QY 155 ----- 158
DB 751 TAGCTTAATGATGCGCAACAGCTTAATGCTCGCGCATACCAATGGTAAACCCCAAGCC 810
QY 159 ----- 168
DB 811 GTAGCTAAAGACCTAGTACAGAGCTAACCCCTGAAGAAAAAACATAGTACAGGGTTA 870
QY 169 IlePheIleThrValMetArgAsnAsp-----GlyLeuSerIleIleSer 183
DB 871 CTAGCTAAGACTATTGGAAGGGGTGAAGTTGTGAGATCAGGGCGGTTTCTTCTACTTC 930
QY 184 ValIleValAsnValCysTYrAspPheSerLeuAsnLeuSerIleSerProTYrIle 203
DB 931 GTAATGCTCAAGCTTGATGATCTTCTTAAGAAAGTTTGAAGTGTCTTCTTATGCT 990
QY 204 CysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIle-----LYs 221
DB 991 TGTGTTGGCTTGGCGGTAACCTTGGGCGGTGGTTGATGGC---CATATCACTCCTAAG 1047
QY 222 PheAlaTYrGlnSerLYeGlyIleAlaTYrSerLeuProSerAsnIleSerLeuPhe 241
DB 1048 CTTCCTTATAGATTAAAGCTGGGTGAGTATCAGCTTCTTCCCTATATCTCCGCTTT 1107
QY 242 AlaSerLeuTYrTYrHisLYeValMetGlyAsn 252
DB 1108 GCGGCTGAGTCTACCATCGCGTTGGGAGAT 1140

```

RESULT 32  
 US-08-975-762-1  
 ; Sequence 1, Application US/08975762  
 ; Patent No. 6207169

```

; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/975,762
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.439
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1345 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Ehrlichia
; US-08-975-762-1
;
; Alignment Scores:
; Pred. No.: 1,32e-14 Length: 1345
; Score: 196.00 Matches: 68
; Percent Similarity: 41.95% Conservative: 44
; Best Local Similarity: 25.47% Mismatches: 71
; Query Match: 13.10% Indels: 84
; DB: 4 Gaps: 12
;
; US-10-062-624-40 (1-293) x US-08-975-762-1 (1-1345)
;
; QY 41 IleSerGlyGlnTYrLYePheSerValSerValPheSerAsnPhenSerValIleGluThr 60
; DB 387 ATTAAGGGTGGGCATTATAGCAGTCACAGCTATTTCCGAATTTT---GTCAAGACACA 443
; QY 61 AsnValIleThrLYeAsnLeuIleAlaLeuLYsAspValAspSerIleGluThrLYs 80
; DB 443 ----- 443
; QY 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPhenThrIleProTYrThrAlaVal 100
; DB 443 ----- 443
; QY 101 PheGlnAspAsnSerValAsnPhenAsnGlyThrIleGlyTYrThrPheAlaGluGlyThr 120
; DB 444 ----- 485
; QY 121 ArgValGluIleGlyLYeSerTYrGluIlePheAspValLYsAsnProGlyLYrTYrThr 140
; DB 486 ----- 506
; QY 141 LeuSerAspAlaTYrArgTYrPheAlaLeuAlaArgGluMetLYeGlyAsnSerPheThr 160
; DB 507 ----- 548

```

Qy 161 Prolysgluys-----ValSerAsnSerIlePhe 170  
Db 549 CCTGAAGAACGACCATAGTGGCTGGTTACTGCTAAATATTATGAAGAGC----- 602  
Qy 171 HisThrValMetArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysThr 190  
Db 603 ---GAGTTATTGAGATTAGGGCCATCTCTCGACTTCACTTACAAATGAAATTTGGTCA 659  
Qy 191 AspPheSerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAsp 210  
Db 660 GATATCAGCATTAATATATC---TTAATGCCGATGTTGGTGGTCCAGGAGTGAAC 716  
Qy 211 AlaIleGluPhePheAsp---ValLeuHisIleLeuPheAlaTyrGlnSerIleGlu 229  
Db 717 TTCTGTTAGTGTGTTGATGATGTCACACTGTCGAAAGTTGCTATTCGTTAAGCAGCT 776  
Qy 230 IleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLeuVal 249  
Db 777 CTGAGTTATAAATTTTCGAAGAGTTACAGCTTTTGACGGTGGTTTACCATCAGCT 836  
Qy 250 MetGlyAsn---GlnPheLeuAsnLeuAsnValGlnHisValAlaGluLeuAlaSerIle 268  
Db 837 ATAGAGATGCTGTTATGATGATCTGCCATTGCGCATTTATCGATGATATTAGTCT 896  
Qy 269 ProlysiLeuThr-----SerAlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGlu 286  
Db 897 GTGAACATGCTTAAGAAACCCCATTCCTGATTCGATGATGAGTACTTGGCGGAGAA 956  
Qy 287 IleGlyAlaArgLeuThrPhe 293  
Db 957 TTGGTGTAGGCTCGCTTT 977

## RESULT 33

US-08-821-324-1  
Sequence 1, Application US/08821324  
Patent No. 6231869  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Houghton, Raymond  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESS: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/821,324  
FILING DATE: 21-MAR-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Makl, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.439  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-682-4900  
TELEFAX: 206-682-6031  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1345 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:  
ORGANISM: Ehrlichia  
US-08-821-324-1

Alignment Scores:  
Pred. No.: 1,32e-14 Length: 1345  
Score: 196.00 Matches: 68  
Percent Similarity: 41.95% Conservative: 44  
Best Local Similarity: 25.47% Mismatches: 71  
Query Match: 13.10% Indels: 84  
Gaps: 12

US-10-062-624-40 (1-293) x US-08-821-324-1 (1-1345)

Qy 41 IleSerGlyGlnTyrProSerValSerValPheSerAsnPheSerValGluThr 60  
Db 387 ATAAAGGGGTGGCATTTATAGCATCAAGCTATTTCCGAATTTT---GTCAAGACACCA 443  
Qy 61 AsnValIleThrLysAsnLeuIleAlaLeuLysAspValAspSerIleGluThrLys 80  
Db 443 ----- 443  
Qy 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaVal 100  
Db 443 ----- 443  
Qy 101 PheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThr 120  
Db 444 -----CTACAGAGAAATGTATGATGAGACCTGGCCTACATCTACCTGA--- 485  
Qy 121 ArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThr 140  
Db 486 -----GAGAGAGT---GAGAGTACACAC----- 506  
Qy 141 LeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgLueLysGlyAsnSerPheThr 160  
Db 507 -----AACGCATAGCCGTTGCTTAAGACCTAGTA---AATGAACCTTACT 548  
Qy 161 Prolysgluys-----ValSerAsnSerIlePhe 170  
Db 549 CCTGAAGAACGACCATAGTGGCTGGTTACTGCTAAATATTATGAAGAGC----- 602  
Qy 171 HisThrValMetArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysThr 190  
Db 603 ---GAGTTATTGAGATTAGGGCCATCTCTCGACTTCACTTACAAATGAAATTTGGTCA 659  
Qy 191 AspPheSerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAsp 210  
Db 660 GATATCAGCATTAATATATC---TTAATGCCGATGTTGGTGGTCCAGGAGTGAAC 716  
Qy 211 AlaIleGluPhePheAsp---ValLeuHisIleLeuPheAlaTyrGlnSerIleGlu 229  
Db 717 TTCTGTTAGTGTGTTGATGATGTCACACTGTCGAAAGTTGCTATTCGTTAAGCAGCT 776  
Qy 230 IleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLeuVal 249  
Db 777 CTGAGTTATAAATTTTCGAAGAGTTACAGCTTTTGACGGTGGTTTACCATCAGCT 836  
Qy 250 MetGlyAsn---GlnPheLeuAsnLeuAsnValGlnHisValAlaGluLeuAlaSerIle 268  
Db 837 ATAGAGATGCTGTTATGATGATCTGCCATTGCGCATTTATCGATGATATTAGTCT 896  
Qy 269 ProlysiLeuThr-----SerAlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGlu 286  
Db 897 GTGAACATGCTTAAGAAACCCCATTCCTGATTCGATGATGAGTACTTGGCGGAGAA 956  
Qy 287 IleGlyAlaArgLeuThrPhe 293  
Db 957 TTGGTGTAGGCTCGCTTT 977

## RESULT 34

US-09-295-028-1  
Sequence 1, Application US/09295028

Patent No. 6277381  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Houghton, Raymond L.  
APPLICANT: McNeill, Patricia D.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
TITLE OF INVENTION: AND TREATMENT OF EHRlichia INFECTION  
FILE REFERENCE: 210121.439C4  
CURRENT APPLICATION NUMBER: US/09/295,028  
NUMBER OF SEQ ID NOS: 85  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO: 1  
LENGTH: 1345  
TYPE: DNA  
ORGANISM: Ehrlichia sp.  
US-09-295-028-1

Alignment Scores:  
Pred. No.: 132e-14 Length: 1345  
Score: 196.00 Matches: 68  
Percent Similarity: 41.95% Conservative: 44  
Best Local Similarity: 25.47% Mismatches: 71  
Query Match: 13.10% Indels: 84  
Gaps: 12  
DB: 4

US-10-062-624-40 (1-293) x US-09-295-028-1 (1-1345)

```
QY 41 IleserGIglnTYrlyPserValserValpheserAnpheserVallysgluThr 60
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 387 ATACGGGTGGGCGATATATAGCAGTCACAGCGATTTCGGAATTTT---CTCAAGACACA 443
QY 61 AenValIleThrLysAenLeuIleAlaLeuLysAenValAenSerIlegluThrlys 80
    ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 443 ----- 443
QY 81 ThrAspAlaSerValglYlIleSerAnPserSerAnPheThrIleProtyrThraAlaVal 100
    ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 443 ----- 443
QY 101 PheGlnAspAnSerValAenPheAenGlyThrIleGlyTyrrThrPheAlaGluGlyThr 120
    ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 444 -----CTACAGAGAAATGCTAGAGAACTGGCCATCATCTACTGGA--- 485
QY 121 ArgValIleGluIleGluIleGlySerTyrrGluGluPheAenValLysAnPProGlyGlyTyrrThr 140
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 486 -----GAGAGAGAGT---GAGAGTAAAGAC----- 506
QY 141 LeuSerAspAlaTyrrGlyrrPheAlaLeuAlaArgLumetLysGlyLysAnSerPheThr 160
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 507 -----AAGCCCATAGCCGTGCTAAGAGCTTGA---AATGACTTACT 548
QY 161 ProLysGluLys-----ValSerAnSerIlePhe 170
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 549 CCGAAGAGACGAACCATAGTGGCTGGCTTACTTGTAAATTAATTAAGAGAC----- 602
QY 171 HisThValMetArgAnSerArgLysSerIleIleSerValIleValAnSerIleCysTyrr 190
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 603 ---GAGGTATTGAGATTAGGGCCATCTTCGACTTCACTTACATGATGAATTTGGCTCA 659
QY 191 AenPheSerLeuAnSerAnSerIleSerProtyrrIleCysGlyGlyValaGlyValaP 210
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 660 GATATACCATAGTAATATC---TTAATGCCGTATGTTGCTTGGTCCAGGATGAGC 716
QY 211 AlaIleGluPheAenPhe---ValLeuHsIleLysPheAlaTyrrGlnSerIleGly 229
    ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 717 TTGTAGAGTGTGTGATGCTACACTGCTGCAAGATTTCATATCGTTAAAGCAGGT 776
QY 230 IleAlaTyrrSerLeuProSerAnSerIleSerPheAlaSerLeuTyrrIleLysVal 249
    ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 777 CTGAGTTAATAATTTTCGAAGAAGTTACAGCTTTTGAGGTGTTTACATCATCGTT 836
```

```
QY 250 MetGlyAen---GlnPheLysAenLeuAenValGlnHisValaGluLeuAlaSerIle 268
    ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 837 ATAGAGATGCTGTTTATGATGATCTCCATTCGGCATTTATCTGATATAGTCT 896
QY 269 ProLysIleThr-----SerAlaValAlaThrLeuAnSerIleGlyTyrrPheGlyGly 286
    ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 897 GTGAAGATGCTAAGAAACCGCATTCATGATTCCTCATAGTACTTTGGCGGGGAA 956
QY 287 IleGlyAlaArgLeuThrPhe 293
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 957 TTGTGTTAGGCTCGCTTT 977
```

RESULT 35  
US-09-106-582-1

Sequence 1, Application US/09106582

Patent No. 6306402

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Lodes, Michael J.

APPLICANT: Houghton, Raymond

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND

NUMBER OF SEQUENCES: 73

CORRESPONDENCE ADDRESS:

ADDRESS: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/106,582

FILING DATE: 29-JUN-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Mark, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.439C2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-622-4900

TELEFAX: 206-682-6031

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1345 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Ehrlichia

US-09-106-582-1

Alignment Scores:

Pred. No.: 1.32e-14 Length: 1345  
Score: 196.00 Matches: 68  
Percent Similarity: 41.95% Conservative: 44  
Best Local Similarity: 25.47% Mismatches: 71  
Query Match: 13.10% Indels: 84  
Gaps: 12  
DB: 4

US-10-062-624-40 (1-293) x US-09-106-582-1 (1-1345)

```
QY 41 IleserGIglnTYrlyPserValserValpheserAnpheserVallysgluThr 60
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 387 ATACGGGTGGGCGATATATAGCAGTCACAGCGATTTCGGAATTTT---CTCAAGACACA 443
QY 61 AenValIleThrLysAenLeuIleAlaLeuLysAenValAenSerIlegluThrlys 80
    ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 443 ----- 443
```





Db 1849 TATGATGATCTGCATTCGGCATTTATCTGATGATATTAAGTCTGTGAACATCTAG 1908  
QY 273 ---SerAlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyValIleGlyAlaAspLeu 291  
Db 1909 GAACCGCATTCATGATTCCTCATAGAGTACTTGGCGGGAAATTGGTTAGGCTC 1968  
QY 292 ThrPhe 293  
Db 1969 GCTTTT 1974

RESULT 37  
US-08-975-762-40  
; Sequence 40, Application US/08975762  
; Patent No. 6207169  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND  
; NUMBER OF SEQUENCES: 73  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: 21-MAR-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Marki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.439  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-622-4900  
TELEFAX: 206-682-6031  
INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1919 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-975-762-40

Alignment Scores:  
Pred. No.: 2,976-13 Length: 1919  
Score: 187.00 Matches: 50  
Percent Similarity: 48.99% Conservative: 23  
Best Local Similarity: 33.56% Mismatches: 56  
Query Match: 12.50% Indels: 20  
DB: 4 Gaps: 5

US-10-062-624-40 (1-293) x US-08-975-762-40 (1-1919)

QY 150 LeuAlaArgGluMetLeuGlyAsnSerPheThrProGlyGlyValSerAsnSerIle 169  
Db 360 CTAGCTAAACATATTGAAGGGGTGAAGTTGTGAGATCAGGGCGGTTCTTCTACT--- 416  
QY 170 PheHisThrValMetArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCys 189  
Db 417 -----TCTGTGATGGTTAATGCTTGT 437

QY 190 TyrAspPheSerLeuAsnLeuSerIleSerProTyrIleCysGlyGlyValIle 209  
Db 438 TATGATCTTCTTATGAGGTTTAAAGGCTTAAAGGCTTCTTAAAGGCTTCTCGAGGT 497

QY 210 AspaIalIeGluPhePheAspValLeuHisIle-----LysPheAlaTyrGlnSerLys 227  
Db 498 AACTTCGTGGCGCGTTGTATAGG---CATATCACTCTCAAGCTGCTTATAGATTAAG 554  
QY 228 LeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyHis 247  
Db 555 GCTGGCTTGAAGTTATCAAGCTCTCTCTGAAATCTCTCTTTCGCGGGGTTTCTACCAT 614  
QY 248 LysValMetGlyAsn---GlnPheLysAsnLeuAsnValGlnHisValAlaGluLeuAla 266  
Db 615 CGTGTGGAGAGATGGTGTATTATGATGATCTGCCACTCAAGCTTGTGTGATGATACT 674  
QY 267 SerIleProLysIleThr-----SerAlaValAlaThrLeuAsnIleGlyTyrPheGly 284  
Db 675 AGTCGGCGCGCGCTACTAAGATCAAGTCTGTCTCACTTCTCCATGAGCTTATGTCGGT 734  
QY 285 GlyIleIleGlyAlaArgLeuThrPhe 293  
Db 735 GGGGAATTGGTGTATGAGTTGCTTTT 761

RESULT 38  
US-09-295-028-40  
; Sequence 40, Application US/09295028  
; Patent No. 6277381  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: McNellis, Patricia D.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; FILE REFERENCE: 210121.439C4  
; CURRENT APPLICATION NUMBER: US/09/295,028  
; CURRENT FILING DATE: 1999-04-20  
; NUMBER OF SEQ ID NOS: 85  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 40  
; LENGTH: 1919  
; TYPE: DNA  
; ORGANISM: Ehrlichia sp.  
US-09-295-028-40

Alignment Scores:  
Pred. No.: 2,976-13 Length: 1919  
Score: 187.00 Matches: 50  
Percent Similarity: 48.99% Conservative: 23  
Best Local Similarity: 33.56% Mismatches: 56  
Query Match: 12.50% Indels: 20  
DB: 4 Gaps: 5

US-10-062-624-40 (1-293) x US-09-295-028-40 (1-1919)

QY 150 LeuAlaArgGluMetLysGlyAsnSerPheThrProGlyGlyValSerAsnSerIle 169  
Db 360 CTAGCTAAACATATTGAAGGGGTGAAGTTGTGAGATCAGGGCGGTTCTTCTACT--- 416  
QY 170 PheHisThrValMetArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCys 189  
Db 417 -----TCTGTGATGGTTAATGCTTGT 437

QY 190 TyrAspPheSerLeuAsnLeuSerIleSerProTyrIleCysGlyGlyValIle 209  
Db 438 TATGATCTTCTTATGAGGTTTAAAGGCTTAAAGGCTTCTTAAAGGCTTCTCGAGGT 497

QY 210 AspaIalIeGluPhePheAspValLeuHisIle-----LysPheAlaTyrGlnSerLys 227  
Db 498 AACTTCGTGGCGCGTTGTATAGG---CATATCACTCTCAAGCTGCTTATAGATTAAG 554  
QY 228 LeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyHis 247  
Db 555 GCTGGCTTGAAGTTATCAAGCTCTCTCTGAAATCTCTCTTTCGCGGGGTTTCTACCAT 614

Qy 248 LysValMetGlyAsn---GlnPheLysAsnLeuAsnValGlnHisValAlaGluLeuAla 266  
Db 615 CGGTGTGGGAGAGAGGTTATGATGATCTGCCAGCTCAACGCTTGTATGATGATCT 674  
Qy 267 SerLeuProLysIleThr-----SerAlaValAlaThrLeuAsnIleGlyTyrPheGly 284  
Db 675 AGTCGGGGGGGCGGCTGACTAGAGATCTGCTGTGCTACTCTCATGCTGATGCTCGGT 734  
Qy 285 GlyGluIleGlyAlaArgLeuThrPhe 293  
Db 735 GGGGAAATTTGGTGTAGGTTTCCTTT 761

RESULT 39  
US-09-106-582-40  
Sequence 40, Application US/09106582  
Patent No. 6306402  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Houghton, Raymond  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/106,582  
FILING DATE: 29-JUN-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Mark, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121,439C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-622-4900  
TELEFAX: 206-682-6031  
INFORMATION FOR SEQ ID NO: 40:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1919 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-106-582-40

Alignment Scores:  
Pred. No.: 2,976-13 Length: 1919  
Score: 187.00 Matches: 50  
Percent Similarity: 48.99% Conservative: 23  
Best Local Similarity: 33.56% Mismatches: 56  
Query Match: 12.50% Indels: 20  
DB: 4 Gaps: 5

US-10-062-624-40 (1-293) x US-09-106-582-40 (1-1919)

Qy 150 LeuAlaArgGluMetLysGlyAsnSerPheThrProLysGluLysValSerAsnSerIle 169  
Db 360 CTAGCTAAACATATTGAAGGCGGTGAGATGATCAGCGCGCTTCTTCTACT--- 416  
Qy 170 PheHisThrValMetArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCys 189  
Db 417 -----TCTGTATGAGTTTATCTTGT 437  
Qy 190 TyrAspPheSerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyValAlaGlyVal 209

Db 438 TATGATCTCTTAGTGAAGGTTTAGCGCTTGTCTTACGGCTTGTGCTGCTGAGAGT 497  
Qy 210 AspAlaIleGluPheAspValLeuHisIle-----LysPheAlaTyrGlnSerIys 227  
Db 498 AACTTCGGGGCGGTGTGATGAGG---CATATCACTCTTAAGCTTCTTAAGATTAAG 554  
Qy 228 LeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrTrpHis 247  
Db 555 GCTGCTGAGTTATCAGCTCTCTCTCAAACTCTGCTTTCCTTGGGGGTTTCAACAT 614  
Qy 248 LysValMetGlyAsn---GlnPheLysAsnLeuAsnValGlnHisValAlaGluLeuAla 266  
Db 615 CGGTGTGGGAGAGAGGTTATGATGATCTGCCAGCTCAACGCTTGTATGATGATCT 674  
Qy 267 SerLeuProLysIleThr-----SerAlaValAlaThrLeuAsnIleGlyTyrPheGly 284  
Db 675 AGTCGGGGGGGCGGCTGACTAGAGATCTGCTGTGCTACTCTCATGCTGATGCTCGGT 734  
Qy 285 GlyGluIleGlyAlaArgLeuThrPhe 293  
Db 735 GGGGAAATTTGGTGTAGGTTTCCTTT 761

RESULT 40  
US-08-953-326-13  
Sequence 13, Application US/08953326  
Patent No. 6251872  
GENERAL INFORMATION:  
APPLICANT: Bardet, Anthony F.  
APPLICANT: Ganta, Roman R.  
APPLICANT: McGuire, Travis C.  
APPLICANT: Burridge, Michael J.  
APPLICANT: Nyika, Aceme  
APPLICANT: Rurangirwa, Fred R.  
APPLICANT: Mahan, Suman M.  
TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of  
TITLE OF INVENTION: Animals and Humans  
FILE REFERENCE: UF-167C1  
CURRENT APPLICATION NUMBER: US/08/953,326  
CURRENT FILING DATE: 1997-10-17  
EARLIER APPLICATION NUMBER: 08/953,326  
EARLIER FILING DATE: 1997-10-17  
EARLIER APPLICATION NUMBER: 08/733,230  
EARLIER FILING DATE: 1996-10-17  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 13  
LENGTH: 399  
TYPE: DNA  
ORGANISM: Ehrlichia canis  
US-08-953-326-13

Alignment Scores:  
Pred. No.: 1,416-12 Length: 399  
Score: 173.00 Matches: 53  
Percent Similarity: 51.08% Conservative: 18  
Best Local Similarity: 38.13% Mismatches: 58  
Query Match: 11.56% Indels: 10  
DB: 4 Gaps: 6

US-10-062-624-40 (1-293) x US-08-953-326-13 (1-399)

Qy 1 MetAsnAsnLysLeuLysPheThrIleIleAsnThrValLeuValCysLeuLeuSerLeu 20  
Db 3 ATGATTTGTAATAAAAGTTTACACA---ATAAGTCATGATGATATCATCATATCTTCTCTA 59  
Qy 21 ProAsnIleSerSerSerLysAlaIleAsnAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40  
Db 60 CTTATGCTCTATCTACTTACCAACCAATATATGTAACAGT---ATGATATGTAATTTTAC 116  
Qy 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnAsnSerValLysGluThr 60  
Db 117 ATATCAGAAAGTACATCCCAAGTGTTCCTCATTTTGAATTTTTCAGCTGAAGAGAG 176

```
Qy 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
Db 177 AAAAAAAAAACACTGTAGTATATGCTTAAAGAAAC-----TGGGCA 221
Qy 81 ThrAspAlaSerValGlyLysSerAsnProSerAsnPheThrIle--ProTyrThrAla 99
Db 222 GGAGATGCATAATCTAGTCAAAGTCCAGATGATTAATTACCATTCGAAATTACTCATTC 281
Qy 100 ValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGlu 118
Db 282 AAGTATGCAGCAACAGATTATTAGGTTTGACAGTAGCTATTGGTTACTCGATAGGC--- 338
Qy 119 GlyThrArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGly 137
Db 339 AGTCCAGAAATAGAAAGTGTGATGCTTTATGAAGCATTTGATGTAAAAAATCAAGGT 395
```

Search completed: July 3, 2003, 23:31:45  
Job time : 67 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 3, 2003, 22:34:29 ; Search time 191 Seconds

(without alignments)  
3454.636 Million cell updates/sec

Title: US-10-062-624-40

Perfect score: 1496

Sequence: 1 MNNLKFTIITVLCILSL.....AVATLNGVFGSEIGARLTF 293

Scoring table:

BLOSUM62	Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```
-MODL=frame_plus_p2n.model -DEV=xlh
-Q=/cgn2.1/USPFO.epool/US10062624/runat_30062003_091228_24295/app_query.fasta.1.455
-DB=Geneseq.101002 -QFMT=fasta -SUFFIX=ing -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40
-MODE=LOCAL -OUTPM=pcio -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10062624.GCGN.1.1.200 @runat_30062003_091228_24295 -NCPU=6 -ICPU=3
-NO_MMAP -LARGESOURCE -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database : N\_Geneseq.101002.\*

```
1: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1496	100.0	879	24	ABK68875
2	1496	100.0	882	20	ABK68875
3	1026	68.6	894	20	ABK68875
4	1008.5	67.4	891	20	ABK68875
5	930	62.2	817	20	ABK68875
6	923	61.7	900	20	ABK68875
7	789	52.6	858	20	ABK68875
8	786.5	52.6	840	20	ABK68875
9	768	51.3	852	20	ABK68875
10	737.5	49.3	828	20	ABK68875
11	732	48.9	882	20	ABK68875
12	671	44.9	888	20	ABK68875
13	574	38.4	493	20	ABK68875
14	500.5	33.5	488	20	ABK68875
15	498	33.3	852	20	ABK68875
16	490.5	32.8	840	24	ABK68875
17	470.5	31.5	843	20	ABK68875
18	470.5	31.5	867	20	ABK68875
19	468.5	31.3	843	21	ABK68875
20	468.5	31.3	4683	19	ABK68875
21	468.5	31.3	4683	21	ABK68875
22	468.5	31.3	4683	22	ABK68875
23	459	30.7	864	19	ABK68875
24	459	30.7	864	22	ABK68875
25	459	30.7	864	22	ABK68875
26	450.5	30.1	843	20	ABK68875
27	449.5	30.0	2037	21	ABK68875
28	446.5	29.8	840	21	ABK68875
29	446.5	29.8	840	24	ABK68875
30	445.5	29.8	843	20	ABK68875
31	443.5	29.6	924	20	ABK68875
32	443.5	29.6	1607	21	ABK68875
33	443.5	29.6	1607	21	ABK68875
34	441.5	29.5	837	20	ABK68875
35	441.5	29.5	837	21	ABK68875
36	441	29.5	849	21	ABK68875
37	441	29.5	849	24	ABK68875
38	440	29.4	852	20	ABK68875
39	437	29.2	813	24	ABK68875
40	434	29.0	846	20	ABK68875
41	431	28.8	1570	21	ABK68875
42	430	28.7	1570	19	ABK68875
43	430	28.7	1570	22	ABK68875
44	429.5	28.7	831	20	ABK68875
45	427	28.5	861	20	ABK68875

## ALIGNMENTS

RESULT 1  
ID ABK68875 standard; DNA; 879 BP.

AC ABK68875;

DT 02-JUL-2002 (first entry)

DE DNA encoding Ehrlichia canis p28-1.

KW Ehrlichia canis infection; vaccine; serodiagnostic; gene; p28; ss;

KW antibacterial.

OS Ehrlichia canis.

XX WO200222782-A2.

XX 21-MAR-2002.



Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 20 Gaps: 0

US-10-062-624-40 (1-293) x AAX34749 (1-894)

```

QY 1 MetAsnAsnLysLeuLysPheThrIleIleAsnThrValLeuValCysLeuLeuSerLeu 20
DB 1 ATGAATATAATACTCAAAATTTACTATATAAACAAGATTAGTATGCTTATTGCTATTA 60
QY 21 ProAsnIleSerSerSerLysAlaIleAsnAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40
DB 61 CCTATATATATCTCTCTCAAGGCCATTAACAATAACGCTAAAGAAAGTACTACGAGTATAT 120
QY 41 IleserGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValIleGluThr 60
DB 121 ATCACTGGACATATATAACCCAGGTTCTGTTTCAGTAATTTTCAGTTAAAGAAACC 180
QY 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
DB 181 AATGTCATATCAATAAAACCTTATACCTTAAAGAAAGTGGACTATATGAACCAAG 240
QY 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaVal 100
DB 241 ACTGATGCCAGTGTAGTATTAGTAAACCATCAATTAATTTACTATCCCTATATACAGCTGTA 300
QY 101 PheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThr 120
DB 301 TTTCAGATTAATATCTGTCATATTCATAGAACATATGTTGTTACACCTTTGCTGAAGTACA 360
QY 121 ArgValGluIleGluGlySerTyrGlyGluPheAspValLysAsnProGlyGlyTyrThr 140
DB 361 AGAGTTGAATATAGAAAGCTTCTTATAGAGATTTGATGTTAAACCCCTGGAGGCTATACA 420
QY 141 LeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgIleLysGlyLysAsnSerPheThr 160
DB 421 CTAAGTGTAGCTCATTCCTATTTGCTATTAGCAGCATGAAGTAAAGTAAATAGTTTACA 480
QY 161 ProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSer 180
DB 481 CCTAAAGAAAAGTTCTTAATAGTATTTTTCACACTGTATAGAAATGATGAGATTATCT 540
QY 181 IleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSer 200
DB 541 ATAATATCTGTTATAGTAAATGTTTGTCTACATTTCTCTTGAACAAATTTGTCAATATCG 600
QY 201 ProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPheAspValLeuHisIle 220
DB 601 CCTATCATATGTGGAGGAGCAGGGGTAGATGCTATAGAAATCTTCGATGATTAACAATT 660
QY 221 LysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeu 240
DB 661 AAGTTTGCATATCAAGCAAGCTAGTATGCTTATTTCTTACCACTCAATCAATTAATGCTTC 720
QY 241 PheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGln 260
DB 721 TTTCGATATATATATCAATTAAGTAAATGGGCAATCAATTTAAATTAATGCTCAAA 780
QY 261 HisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIle 280
DB 781 CATGTGTGTAAGCTGCAAGTATATACCTAAATTAATCATCGCAGTTCACCTTAATATT 840
QY 281 GlyTyrPheGlyGlyGluIleGlyValaArgLeuThrPhe 293
DB 841 GGTAATTTTGGAGGTGAATTTGTCAGATTAATGCAATTT 879

```

# RESULT 3.

AAX34749 standard; DNA; 894 BP.

AAX34749;

DT 05-JUL-1999 (first entry)

```

XX XX DNA encoding OMP-1A protein.
DE Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
XX KM detection; dog; ss.
XX OS Ehrlichia chaffeensis.
XX PN W09913720-A1.
XX 25-MAR-1999.
XX 18-SEP-1998; 98WC-US19600.
XX 19-SEP-1997; 97US-0059353.
XX (OHIS ) UNIV OHIO STATE.
XX PA
XX PI Ohashi N, Rikihisa Y;
XX DR WPI: 1999-254290/21.
XX P-PSDB; AAY06949.
XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and
XX PT Ehrlichia canis
XX PS
XX SS Disclosure; Fig 9A; 55pp; English.
XX CC The invention provides isolated outer membrane proteins (OMP) from
XX CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
XX CC of the OMP family and consist of proteins OMP-1, -1B to -Z) shown
XX CC in AAY06943-958. The E. canis proteins form part of the P30 family and
XX CC consist of proteins shown in AAY06959-970. The proteins and genes are
XX CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX SQ Sequence 894 BP; 308 A; 143 C; 152 G; 291 T; 0 other;

Alignment Scores:
Pred. No.: 7.64e-102 Length: 894
Score: 1026.00 Matches: 194
Percent Similarity: 79.59% Conservative: 40
Best Local Similarity: 65.99% Mismatches: 58
Query Match: 68.58% Indels: 2
DB: 20 Gaps: 2

US-10-062-624-40 (1-293) x AAX34749 (1-894)
QY 1 MetAsnAsnLysLeuLysPheThrIleIleAsnThrValLeuValCysLeuLeuSerLeu 20
DB 1 ATGAATATAATACTCAAAATTTACTATATAAACAAGATTAGTATGCTTATTGCTATTA 72
QY 21 ProAsnIleSerSerSerLysAlaIleAsnAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40
DB 73 CCTGATATATCATTTTCAAGAACTATTAACAACAGTCTAAAGAAAGCTGGGTTATAT 132
QY 41 IleserGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValIleGluThr 60
DB 133 ATCAGTGGCAGTACCAACCTAGGTTTTCAGTTTATGTAATTTTTCAGTAAAGAAACT 192
QY 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
DB 193 AATGTTCCACAAAGCAGTTTATATGCACTTAAAGAAAGTATATTTCTGTTGCGATTGGT 252
QY 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaVal 100
DB 253 AGTATGCTACTACAGGATATAGCAATCCAGGTATTTCACAATTCCTTATACTGACGAA 312
QY 101 PheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThr 120
DB 313 TTTCAGATTAATATGTCATTAATTTCAATGAGGCTTGTGTTACTCTTTCCGATATGCTA 372
QY 121 ArgValGluIleGluGlySerTyrGlyGluPheAspValLysAsnProGlyGlyTyrThr 140
DB 121 ArgValGluIleGluGlySerTyrGlyGluPheAspValLysAsnProGlyGlyTyrThr 140

```

Db 373 AGAATTGAATAGAGGATTCATGATAAATTGATGCAAAAACCTGAGGTTACACA 432  
 Qy 141 ---LeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgLysMetLeuGlyAsnSerPhe 159  
 Db 433 CAAATAAAGATGCGATTCATTTGATGACATGACGATTAAGATTAAGATGCTTCTT 492  
 Qy 160 ThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeu 179  
 Db 493 GAACTTAAGCGGAA---GATACAGGTGTTTATCATCTGATTAAGAAAATGATGATTA 549  
 Qy 180 SerIleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIle 199  
 Db 550 TCTATTATTAATCTACATGCTTACGCTGTTACGATTTCTGATGATGAATTAACCACTC 609  
 Qy 200 SerProTyrIleCysGlyGlyValAlaGlyValAspAlaIleGluPhePheAspValLeuHis 219  
 Db 610 TTACCTTATATATGTCAGAGTATGGGTTAAAGCCATTAAGATTTCTTGACCTTTACAT 669  
 Qy 220 IleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSer 239  
 Db 670 GTAAATTTTGTCTTACCAAGGCAAGCAAGTATTAAGTATTAAGTAAAT 729  
 Qy 240 LeuPheAlaSerLeuTyrTyrHisValMetGlyAsnGlnPheLysAsnLeuAsnVal 259  
 Db 730 TTATTCCTTGATGGGATTAATCAATCAAGTAATAGCAATCAATCAAAACTTAACGTA 789  
 Qy 260 GlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsn 279  
 Db 790 AACCAATGTTACACACTTAAGAAATCTCTAAAGTCACATCTGACAGTACACTGAC 849  
 Qy 280 IleGlyTyrPheGlyGlyIleGlyValAlaArgLeuThrPhe 293  
 Db 850 ATTGCATCTTTGGTGGCGAAGTGGAAATTAATTAACATTT 891  
 RESULT 4  
 AAX34767  
 ID AAX34767 standard; DNA; 891 BP.  
 AC AAX34767;  
 XX 05-JUL-1999 (first entry)  
 DT 05-JUL-1999  
 DE DNA encoding P30-7 protein.  
 XX  
 KM Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 detection; dog; se.  
 OS Ehrlichia canis.  
 ON  
 PN MO9913720-AL.  
 XX 25-MAR-1999.  
 PD 18-SEP-1998; 98WO-US19600.  
 PF 19-SEP-1997; 97US-0059353.  
 PR  
 XX (OHIS ) UNITV OHIO STATE.  
 PA  
 XX Ohashi N, Rikihisa Y;  
 PI WPI; 1999-254290/21.  
 DR P-PSDB; AAY06967.  
 XX Novel outer membrane proteins from Ehrlichia chaffeensis and  
 PT Ehrlichia canis  
 PS Disclosure; Fig 27A; 55pp; English.  
 CC The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and

CC consist of proteins shown in AAY06959-970. The proteins and genes are  
 CC used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX  
 SQ Sequence 891 BP; 327 A; 135 C; 130 G; 299 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 6.04e-100 Length: 891  
 Score: 1008.50 Matches: 200  
 Percent Similarity: 80.00% Conservative: 36  
 Best Local Similarity: 67.80% Mismatches: 54  
 Query Match: 67.41% Indels: 5  
 DB: 20 Gaps: 4  
 US-10-062-624-40 (1-293) x AAX34767 (1-891)  
 Qy 1 MetAsnLysLeuLysPheThrIleIleAsnThrValLeuValCysLeuLeuSerLeu 20  
 Db 13 ATGAATATATTAAGCTATCTTATATAGATTATATTTATTAACATGATCTGTCATTA 72  
 Qy 21 ProAsnIleSerSerSerLysAlaIleAsnAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40  
 Db 73 CCTATATATCTCTTCAAA---GTAAATAC-----GAAATAACATTCGTTGTAT 123  
 Qy 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60  
 Db 124 ATTAGCGGGCAATACAAACCAGAGTTCGTTTCAGTAAATTTTCAGTTAAAGAAACC 183  
 Qy 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80  
 Db 184 AACTTCATACAAACATCTCATAGCTTTAAAGAGTGTGATCTGTGAATGAT 243  
 Qy 81 ThrAspAla---SerValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAla 99  
 Db 244 ACTGTAATATATACAGCAGGATTAATGTAACCATTAACCTTATACATCCCTTATACGCA 303  
 Qy 100 ValPheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGly 119  
 Db 304 GAATTTCAAGACACACATCACTCAATGCAATGCTCTATGTTGTTGCTTGTGAAGGT 363  
 Qy 120 ThrArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyr 139  
 Db 364 CCAAGAAATGAATGAATTAATCATATGAATAATTAAGTTAAATCCACAGGGGAT 423  
 Qy 140 ---ThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgLysMetLeuGlyAsnSer 158  
 Db 424 ACTACAGTAAAGATGCTTAATGATCTTCTTACGACCTGAATTAATATTTCTCTA 483  
 Qy 159 PheThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGly 178  
 Db 484 TTCACACCAAAACAAAAGAGTAGTGAATTAACATGCGTAATGAAAAACGATGG 543  
 Qy 179 LeuSerIleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSer 198  
 Db 544 TTATCTATCTTATCAATATATGTAATATGTTACGATTTCTTAAATTAATTTACTT 603  
 Qy 199 IleSerProTyrIleCysGlyGlyValAlaGlyValAspAlaIleGluPhePheAspValLeu 218  
 Db 604 ATATCACTTATTTATGCGAGGAGATGGGTAAATGCAATTAATTTCTTGACCTTTA 663  
 Qy 219 HisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIle 238  
 Db 664 CATGGAATTTGCTTATCAAGCAGGCAAGAAATGATTAACAATTAACGTAAATTC 723  
 Qy 239 SerLeuPheAlaSerLeuTyrTyrHisValMetGlyAsnGlnPheLysAsnLeuAsn 258  
 Db 724 AACTTAATTAATGATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 783  
 Qy 259 ValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeu 278  
 Db 784 GTCCAACTGTACATGAATTAAGATTAATCAAAAGTCACATCTGACGTGTCACATTT 843  
 Qy 279 AsnIleGlyTyrPheGlyGlyIleGlyValAlaArgLeuThrPhe 293



Db 844 GATATGACATATTTGTAGTGAAGCTGGCATTAAGATTAATTT 888  
 RESULT 5  
 AAX34758  
 ID AAX34758 standard; DNA; 817 BP.  
 XX  
 AC AAX34758;  
 XX  
 DT 05-JUL-1999 (first entry)  
 XX  
 DE DNA encoding OMP-12 protein.  
 XX  
 KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 KW detection; dog; ss.  
 XX  
 OS Ehrlichia chaffeensis.  
 XX  
 PN WO9913720-A1.  
 XX  
 PD 25-MAR-1999.  
 XX  
 PE 18-SEP-1998; 98WO-US19600.  
 XX  
 PR 19-SEP-1997; 97US-0059353.  
 XX  
 PA (OHIS ) UNIV OHIO STATE.  
 XX  
 PI Ohashi N, Rikihisa Y;  
 XX  
 DR WPI, 1999-254290/21.  
 DR P-PSDB; AAY06958.  
 XX  
 PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
 PT Ehrlichia canis  
 XX  
 PS Disclosure; Fig 18A; 55pp; English.  
 XX  
 CC The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and  
 CC consist of proteins shown in AAY06959-970. The proteins and genes are  
 CC used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX  
 SQ Sequence 817 BP; 264 A; 132 C; 145 G; 276 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1.79e-91 Length: 817  
 Score: 930.00 Matches: 180  
 Percent Similarity: 79.34% Conservative: 35  
 Best Local Similarity: 66.42% Mismatches: 48  
 Query Match: 62.17% Indels: 8  
 DB: 20 Gaps: 3  
 US-10-062-624-40 (1-293) x AAX34758 (1-817)  
 QY 31 AsnAenAlAlYbLYrTYrGLYrLYrLIEsErGLYrLYrLpProSeRvAlSer 50  
 Db 2 AGCAGCCTTAAGAAAAAGTTGGCTTATATGTAGTGACACACAGCTTGTCTTCT 61  
 QY 51 ValPheSeRvAenPhSeRvAlYbGLYrThraPheValIleThLysE--AsnLeuIleAla 69  
 Db 62 ATTTTTCAGATTTCTCAGTAAAGAACTAAATTTTCTTCTACAAAGTATTTGACGCTTC 121  
 QY 70 LeuLYbLYbAsPvAlAsPSeRvIleGLYrThraPheValIleThLysE--AsnLeuIleAla 89  
 Db 122 TTAAGAAAAAGACATTAATCTGTGCAATTGACGATGCTTACGCGCATTTAGTTAC 181  
 QY 90 ProSeRvAenPhThrIleProTYrThraValAlPheGLInAsPSeRvAlAsnPhenAln 109  
 Db 182 CCACTTAATTTTCAGTACTCTTATATAGCTGATTTCAAGATAATTTCTTAATTTTAAAT 241  
 QY 110 GlyThrIleGLYrThrPheAlaGLYrThraGValGLuIleGLuGLYrSerTYrGLu 129

Db 242 GGCGCTATTGGGTACACTTTTGTGAAGGCCCAAGAAATGAAGAGTTCTTAAGAA 301  
 QY 130 GluPheAsPvAlYbAsnProGLYrLYrThr---LeuSeRvAlAlTYrArTYrPhe 148  
 Db 302 GAATTCATGTCAAAGACCTCGAAGATATACAAATATACAGATGCAATACGTTACTTT 361  
 QY 149 AlAlenAlaArgGLuMetLYbLYbAsnSerPheThrProLYbGLYrValSerAsn--- 167  
 Db 362 GCTTTAGCAGCTGATATAGACTCTATTCCCTACTAGCCCAAAAATAGAACTTCACTGAT 421  
 QY 168 -----SerIlePheIstThrValMetArgAsnAsPGLYbLeuSerIle 182  
 Db 422 GGCAACAGTTTCATATTAAGGTATACCACTGATGAAGAAATGAAGACTATATATA 481  
 QY 183 SerValIleValAlnValCYrTYrAsPSeRvIleAsnLeuSerIleSeRvProTYr 202  
 Db 482 TCCATTAATGGCAATGGCTGCTATGATGATTTTCTTCAGATTAATTAATTAATTAAT 541  
 QY 203 IleCYbGLYrValAGLYbValAsPAlAlleGLuPheAsPvAlLeuHisIleLYbPhe 222  
 Db 542 GTATGTGGTGTATAGGTGTAAATGCTATAGAGTTTTCATGCAATTACATGTTAAATTC 601  
 QY 223 AlaTYrGLuSerLYbLeuGLYrIleAlaTYrSerLeuProSeRvAlSeRvLeuPheAla 242  
 Db 602 GCGTGTACGGGTAAATTAAGTATTACTTATTCATTAATCTTCAACGTTAGTTATTGCT 661  
 QY 243 SerLeuTYrTYrHisLYbValMetGLYbAsnGLuPheAsnLeuAlnValAlnHisVal 262  
 Db 662 GGTGATATTAATTCACCAAGTATGCGCAACCAATTTAAATCTAAATGTTCAACATGTA 721  
 QY 263 AlaGLuLeuAlaSerIleProLYbIleThrSerAlaValAlaThrLeuAsnIleGLYr 282  
 Db 722 GCTGAATTAATGACGCCACCAAGTTATCACTCTCAGTAGCTTACACTTGACATTGGGTAT 781  
 QY 283 PheGLYrGLuIleGLYrAlaArgLeuThrPhe 293  
 Db 782 TTTGGTGTGAATTTGGAGCAAGCTTATATTT 814  
 RESULT 6  
 AAX34768  
 ID AAX34768 standard; DNA; 900 BP.  
 XX  
 AC AAX34768;  
 XX  
 DT 05-JUL-1999 (first entry)  
 XX  
 DE DNA encoding P30-8 protein.  
 XX  
 KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 KW detection; dog; ss.  
 XX  
 OS Ehrlichia canis.  
 XX  
 PN WO9913720-A1.  
 XX  
 PD 25-MAR-1999.  
 XX  
 PE 18-SEP-1998; 98WO-US19600.  
 XX  
 PR 19-SEP-1997; 97US-0059353.  
 XX  
 PA (OHIS ) UNIV OHIO STATE.  
 XX  
 PI Ohashi N, Rikihisa Y;  
 XX  
 DR WPI, 1999-254290/21.  
 DR P-PSDB; AAY06958.  
 XX  
 PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
 PT Ehrlichia canis  
 XX  
 PS Disclosure; Fig 28A; 55pp; English.

```

XX  The invention provides isolated outer membrane proteins (OMP) from
CC  Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC  of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC  in AAY06943-958. The E. canis proteins form part of the P30 family and
CC  consist of proteins shown in AAY06959-970. The proteins and genes are
CC  used to detect E. chaffeensis in patients and E. canis in dogs.
XX
SQ  Sequence 900 BP; 324 A; 142 C; 144 G; 290 T; 0 other;

Alignment Scores:
Pred. No.:      1,18e-90      Length:      900
Score:          923.00      Matches:      182
Percent Similarity: 74.58%      Conservative: 41
Best Local Similarity: 60.87%      Mismatches: 68
Query Match:    61.70%      Indels:      8
DB:             20      Gaps:      4

US-10-062-624-40 (1-293) x AAX34768 (1-900)

QY  1 MetAaenLysLeuLysPheThrIleIleAenThrValLeuValCysLeuLeuSerLeu 20
    1 ATGAATAGCAAGAGTAAAGTTCTTTACAAATATGTAATCGTTAATATGCTTATTAATCA 60
    21 ProAenLysSerSerSerLysAlaIleAenAenAenAlaLysLysTyTyrGlyLeuTyr 40
    61 CCTAACACATCTCTCTCAAACTTCATAGGCAATAGTACA---AAACATCTCGATTAATAT 117
    41 IleserGlyGlnTyrLysProSerValSerValPheSerAenPheSerValLysGluThr 60
    118 GTTAGGGGCAATATTAAGCCAGCGCTTCATTTTACAAATTTTCACTAAAGAAACA 177
    61 AenValIleLysAenLeuIleAlaLeuLysLysAenValAspSerIleGluThrLys 80
    178 AATACACATACAGTACAGTTAGTACTCTTAAAGAAAGTTAATTTCTATTAATGAAAC 237
    81 ---ThrAspLysSerValGlyLysSerAenProSerAenPheThrIleProTyrThrAla 99
    238 ATCAGTAATAGTGCTACAGCATTTAGCAAGCAACAAATTTTAATCTTCTTATGTCGA 297
    100 ValPheGlnAenAenSerValAenPheAenGlyThrIleGlyTyrThrPheAlaGluGly 119
    298 GAATTTCAAGAACATGCTTCAACTTCACTGAGCTAATGCTTATTTGAAACA 357
    120 ThrArgValGlnIleGluLysSerTyrGlnGluPheAenValLysAenProGlyGlyTyr 139
    358 CTAAACATTGAAGTTGAAGTTCTTATGAAGATTCGATGCCAAATATCTGATGTTAT 417
    140 ThrLeuSerAenAlaTyrArgTyrPheAlaLeuAlaArgLysLys----- 154
    418 ATTTTAATGATGATTCGCTATTTTGCATTGGCAGCGTAATGGGACAAAGAAAAAT 477
    155 LysGlyAenSerPheThrProLysGlyLysValSer---AenSerIlePheHisThrVal 173
    478 GATTAATTAACATCTTAAGTCTTAAGGAGCATGATATTAAGTAACATTAATACAGAGC 537
    174 MetArgAenAenPheLysLeuSerIleIleAenValIleValAenValCysTyrAenPheSer 193
    538 ATGAGAAATATAGGTTATCTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 597
    194 LeuAenAenLysSerIleSerProTyrIleCysGlyGlyAlaGlyValAenAlaIleGlu 213
    598 CTCATATGATTTATCAATATCACTTATTTTGTACAGAAATAGGTGATGCTTATTAAGA 657
    214 PhePheAenValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSer 233
    658 TTTTGTGATGCACTGATCTTAACCTTGTTCAGAAAGTAATATAGACTTACTTAACCA 717
    234 LeuProSerAenLysSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAenGln 253
    718 TTAATCAGCAACATTAATGTTATTTACAAATGATATTAATTAATTAATTAATTAATTA 777
    254 PheLysAenLysAenValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSer 273

```

```

DB  778 TTTAAAACTTAATAAGTCATATATATAGTAACTTAAGAAGAACCCGAAATTAATCACT 837
QY  274 AlaValAlaThrLeuAenIleGlyTyrPheLysGlyGluIleGlyAlaArgLeuThr 292
    838 GCAGTGTCTACTTCATATGTTGATGTTGAGGTGAATTTGAGTAAGTCACTCA 894

RESULT 7
AAX34757
ID  AAX34757 standard; DNA; 858 BP.
XX
AC  AAX34757;
XX
DT  05-JUL-1999 (first entry)
XX
DE  DNA encoding OMP-1Y protein.
XX
KW  Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
    detection; dog; ss.
XX
OS  Ehrlichia chaffeensis.
XX
PN  W09913720-A1.
XX
PD  25-MAR-1999.
XX
PF  18-SEP-1998; 98WO-US19600.
XX
PR  19-SEP-1997; 97US-0059353.
XX
PA  (OHIS ) UNIV OHIO STATE.
XX
PI  Ohashi N, Rikihisa Y;
XX
DR  WPI, 1999-254290/21.
XX
DR  P-PSDB; AAY06957.
XX
PT  Novel outer membrane proteins from Ehrlichia chaffeensis and
    Ehrlichia canis
XX
PS  Disclosure; Fig 17A; 55BP; English.
XX
CC  The invention provides isolated outer membrane proteins (OMP) from
CC  Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC  of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC  in AAY06943-958. The E. canis proteins form part of the P30 family and
CC  consist of proteins shown in AAY06959-970. The proteins and genes are
CC  used to detect E. chaffeensis in patients and E. canis in dogs.
XX
SQ  Sequence 858 BP; 293 A; 135 C; 158 G; 272 T; 0 other;

Alignment Scores:
Pred. No.:      3.9e-76      Length:      858
Score:          789.00      Matches:      159
Percent Similarity: 70.41%      Conservative: 48
Best Local Similarity: 54.08%      Mismatches: 77
Query Match:    52.74%      Indels:      10
DB:             20      Gaps:      4

US-10-062-624-40 (1-293) x AAX34757 (1-858)

QY  1 MetAaenLysLeuLysPheThrIleIleAen---ThrValLeuValCysLeuLeuSer 19
    1 ATGAATATATAGAAAGATTTTATTAATAGTGCATCACTTACAGCAAGCTTATTAATC 60
    20 LeuProAenLysSerSerSerLysAlaIleAenAenAenAlaLysLysTyTyrGlyLeu 39
    61 ACATCTGAGGCTCTTCTACAGAAATGTAAGTAACTTAATTAATTAATTAATTAATTAAT 120
    40 TyrIleserGlyGlnTyrLysProSerValSerValPheSerAenPheSerValLysGlu 59
    121 TATATCACTGAGCAATATTAAGCCAGAGTTTCTCATTTTACCAAAATTTTCACTGAA 180

```

```

QY 60 ThrAsnValIleThrLysAsnLeuIleAlaLeuLYsValAspSerIleGluThr 79
DB 181 ACCAACTACATACCTACTCACTAGTGGCTTAAAAAGACATCATGCTCAT 234
QY 80 LysThrAspAlaSerValGlyIleSerAsnProSerAsnPhenThrIleProTyrThrAla 99
DB 235 -----GGGACACGTAATATTCACACCTACACAAATTTCACTTCTTACATTCGCA 285
QY 100 ValPheGlnAspAsnSerValAsnPhenAsnGlyThrIleGlyTyrThrPheAlaGluGly 119
DB 286 GAATTTCAAGACAGATGCGCATTAAGTTTCAGTGGGCGCATGTAATCTTGATTCAGAGAAAT 345
QY 120 ThrArgValGluIleGluGlySerTyrGluGluPheAspValIleAsnProGlyGlyTyr 139
DB 346 TTAGAGATTGAAGTAGAGGCTTCTTAATGAGAAATTTGATTAATTCACCAAGA 402
QY 140 ThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetIleGlyAsnSerPhe 159
DB 403 TCTGCTACAGACGATACAGTATTTTTCATACGACGCTATGATGGC----- 453
QY 160 ThrProLYsGluLYsValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeu 179
DB 484 ACTAATTAATCTAGTCTCGATGATACACAGAAATTCACGTGATGAGAAATGACGCGTTA 513
QY 180 SerIleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIle 199
DB 514 TCAATTTTCACTAGTATGATTAATGAGCTTTCATATTTTACATGATGATATACCAAGA 573
QY 200 SerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHis 219
DB 574 GTACCGTATGATGCGCAGGAATAGGAGAGATTCATGATGATTTTATGATTTACAT 633
QY 220 IleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSer 239
DB 634 GTTAAGTTTGCATCAACAGGCAAGGATGATTTAGTTATCTATATCCCTGAAAGTAAGT 693
QY 240 LeuPheAlaSerLeuTyrTyrHisLYsValMetGlyAsnGlnPheLysAsnLeuAsnVal 259
DB 694 TTATTTCTTAACGATATTACATTAAGTAACAGATTAACAGATTTAAAACTTACACGTT 753
QY 260 GlnHsValAlaGluLeuAlaSerIleProLYsIleThrSerAlaValAlaThrLeuAsn 279
DB 754 CAACACGTAAGTATTAAGTACCGCTCCTTAAGTTACACATCTGCAGTTGCTACATCAAT 813
QY 280 IleGlyTyrPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293
DB 814 GTTGGTACTTGTGGCGAAATTTGAGATTAATTAATTTT 855

```

RESULT 8  
AAK34754  
ID AAK34754 standard; DNA; 840 BP.

AC AAK34754;  
DT 05-JUL-1999 (first entry)  
XX DNA encoding OMP-IV protein.  
DE Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
KW detection; dog; ss.  
XX  
XX Ehrlichia chaffeensis.  
OS  
XX  
XX W09913720-A1.  
XX  
XX 25-MAR-1999.  
XX  
XX 18-SEP-1998; 98WO-US19600.  
XX  
XX 19-SEP-1997; 97US-0059353.  
XX  
XX (OHIS ) UNIV OHIO STATE.  
XX

PI Ohaeshi N, Rikihisa Y;  
XX  
XX WPI: 1999-254290/21.  
DR P-PSDB; AAY06954.  
XX  
XX Novel outer membrane proteins from Ehrlichia chaffeensis and  
PT Ehrlichia canis  
XX  
XX  
PS Disclosure; Fig 14A; 55pp; English.

CC The invention provides isolated outer membrane proteins (OMP) from  
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
CC in AAY06943-958. The E. canis proteins form part of the P30 family and  
CC consist of proteins shown in AAY06959-970. The proteins and genes are  
CC used to detect E. chaffeensis in patients and E. canis in dogs.

CC Sequence 840 BP; 279 A; 128 C; 136 G; 297 T; 0 other;

#### Alignment Scores:

Pred. No.:	7,086-76	Length:	840
Score:	786.50	Matches:	160
Percent Similarity:	67.92%	Conservative:	39
Best Local Similarity:	54.61%	Mismatches:	77
Query Match:	52.57%	Indels:	17
DB:	20	Gaps:	4

US-10-062-624-40 (1-293) x AAK34754 (1-840)

```

QY 3 AsnLYsLeuLYsPheThrIleIleAsnThrValLeuValCysLeuLeuSerLeuProAsn 22
DB 4 AGCAAAAAAAGTTTATTACATAGACATGACTGATGATCTTATTCATTTCTTACT 63
QY 23 IleSerSerLYsValIleAsnAsn-----AlaLYsLYsTyrGlyLeuTyr 40
DB 64 ATTGATCTCTTTCAGCTATTAATCATATCATACAGAAATTAACATAGTGTATATAT 123
QY 41 IleSerGlyGlnTyrLYsProSerValSerValPheSerAsnPhenSerValIleGluThr 60
DB 124 ATTACAGGCGCATATAGCCAGAGATATCCATTTTGCAATTTCTCGTAAGAAACT 183
QY 61 AsnValIleThrLYsAsnLeuIleAlaLeuLYsLYsAspValAspSerIleGluThrLYs 80
DB 184 AATGTTATACATTAACACTGATGATATAAAAAAGGCTCTTCTATGATCTTAC 243
QY 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPhenThrIleProTyrThrAlaVal 100
DB 244 ACT-----TATTCAAACTTTCAGGCTCATATATGTTTACA 279
QY 101 PheGlnAspAsnSerValAsnPhenAsnGlyThrIleGlyTyrThrPheAlaGluGlyThr 120
DB 280 TTTCAAGATTAATGCTGCTGATTTTCAAGTGAACCAATTTGATTTCTTACCCGAAAGTCTA 339
QY 121 ArgValGluIleGluGlySerTyrGluGluPheAspValIleAsnProGlyGlyTyrThr 140
DB 340 AGACTTGACTTGAAGTTCTTACGAAATTTGATGTCAAAGATCTTAAGACTACTCA 399
QY 141 LeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLYsGlyAsnSerPheThr 160
DB 400 GCAAAAGATGCTTTAGGTTTTCCTTACACAGT-----AATACGCTACT 447
QY 161 ProLYsGlyLYsValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSer 180
DB 448 -----ACTGTTCTCGATGCTCAAAAATATACGTTATGAAGAAATATGGCTTATCT 498
QY 181 IleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSer 200
DB 499 GTTGATCAATCAATGATCAATGCTGTTATATATCTATCTTTAATTAATTTAGTGTATCA 558
QY 201 ProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIle 220
DB 559 CTTATATATGTCAGGATATGTCGAAGATTTCAATGAAATTTTGTATATCTTGCACACTT 618

```

```

QY 221 LysPheAlaTyGlnSerLysLeuGlyIleAlaTySerLeuProSerAsnIleSerLeu 240
DB 619 AAACCTGCTTACAGAAAGAACTAGTATTAAGTTTACTTCTTCCCTAAGATTATGA 678
QY 241 PheAlaSerLeuTyTyThiAlaValMetGlyAsnGlnPheLysAsnLeuValGln 260
DB 679 TTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 738
QY 261 HisValAlaGluLeuAlaSerLeuProLysIleThiSerAlaValAlaThrLeuAsnIle 280
DB 739 CAGGTGTTTACACTGATGAAATTTCTTAAGCAACTCTGCAAGTACGCTAATGTT 798
QY 281 GlyTyPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293
DB 799 GCTTATTTTGGTGGTGAAGCTGAGTGAAGTTTAACTTT 837

RESULT 9
AAK34755
ID AAK34755 standard; DNA; 852 BP.
XX
AC AAK34755;
XX
XX 05-JUL-1999 (first entry)
XX
DE DNA encoding OMP-1W protein.
XX
KM Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
XX detection; dog; ss.
XX
OS Ehrlichia chaffeensis.
XX
PN MO9913720-A1.
XX
PD 25-MAR-1999.
XX
PF 18-SEP-1998; 98WO-US19600.
XX
PR 19-SEP-1997; 97US-0059353.
XX
PA (OHIS ) UNIV OHIO STATE.
XX
PI Ohashi N, Rikihisa Y;
XX
DR WPI; 1999-254290/21.
DR P-PSDB; AAY06955.
XX
PT Novel outer membrane proteins from Ehrlichia chaffeensis and
XX Ehrlichia canis
XX
PS Disclosure; Fig 15A; 55pp; English.
XX
CC The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1B to Z) shown
CC in AAY06954-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
SQ Sequence 852 BP; 295 A; 115 C; 141 G; 301 T; 0 other;

Alignment Scores:
Pred. No.: 7,37e-74 Length: 852
Score: 768.00 Matches: 161
Percent Similarity: 68.24% Conservative: 41
Best Local Similarity: 54.39% Mismatches: 78
Query Match: 51.34% Indels: 16
DB: Gaps: 6

US-10-062-624-40 (1-293) x AAK34755 (1-852)
QY 1 MetAsnAsnLysLeuLysPheThrIleIleAsnThValLeuValCysLeuLeuSer--- 19
DB 1 ATGAGTGTCTAAAGAAAGCTTTTATTAAGGCTCAGTGTAGTATGTTAGTTCATCATC 60

```

```

QY 20 LeuPro-----AsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyTy 37
DB 61 TTACTTACTAATATCTTTGCTCAACTTAATATATATATATACACT---AAGTCACT 117
QY 38 GlyLeuTyIleSerGlyGlnTyTyLysProSerValSerValPheSerAsnPheSerVal 57
DB 118 GGGCTATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 177
QY 58 LysGluThrAsnValIleThrLysAsnLeuIleLeuLysLysAspValAspSerIle 77
DB 178 AAAGAACTTACTGACACTTAAGAGATTATAGGACTAGCAAAAGATATTAAGTCTAT 237
QY 78 GluThrLysThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTy 97
DB 238 -----ACAGAT-----ATACAAACAATAAATAAATTCACATCTCTAT 276
QY 98 ThrAlaValPheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyThrPheAla 117
DB 277 AACACAAATTTCAAGATTAATGCTGTAGCTTCAGTCACTGCTGCTGCTGCTGCTGCTGCTGCT 336
QY 118 GluGlyThrArgValGluIleGluGlySerTyTyGluGluPheAspValLysAsnProGly 137
DB 337 GACAGTCCAGAGGTTGAGAGTGAATGCTTATAGAAGATTTGACGTTAAATTCCTGCT 396
QY 138 GlyTyThrLeuSerAspAlaTyArgTyPheAlaLeuAlaArgLysGluMetLysGlyAsn 157
DB 397 AATTACGTAAGTAAGTAAGCTTCAGATATTTGCTTACAGAGGAAATGATAT--- 453
QY 158 SerPheThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAsp 177
DB 454 -----CTTCAAAAATATCTCGAAACAATAAGTATGTTATTAAGAACAT 501
QY 178 GlyLeuSerIleIleSerValIleValAsnValCysTyArgPheSerLeuAsnLeu 197
DB 502 GAGCTTATCTGCGATCCATTATATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 561
QY 198 SerIleSerProTyIleCysGlyGlyAlaGlyValAspAlaIleGluPheAspVal 217
DB 562 AAGATATCACTTACATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 621
QY 218 LeuHisIleLysPheAlaTyGlnSerLysLeuGlyIleAlaTySerLeuProSerAsn 237
DB 622 GTAAGTTTAAATTTGCTTATCAAGTGAAGTACGATACGTTATTCCTCTAT 681
QY 238 IleSerLeuPheAlaSerLeuTyTyThiAlaValMetGlyAsnGlnPheLysAsnLeu 257
DB 682 ATGATTATATTTGCTGACGATATTTACCAATAGGTCATAGGAATTAATTAACAATTTA 741
QY 258 AsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThr 277
DB 742 AATGTTCAACAGCTTGTAGCTTACAGTCAATCTTAAGTCTTACCTTTCAGTACGCTACT 801
QY 278 LeuAsnIleGlyTyPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293
DB 802 CTTATGTTGAGTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 849

RESULT 10
AAK34756
ID AAK34756 standard; DNA; 828 BP.
XX
AC AAK34756;
XX
XX 05-JUL-1999 (first entry)
XX
DE DNA encoding OMP-1X protein.
XX
KM Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
XX detection; dog; ss.
XX
OS Ehrlichia chaffeensis.
XX
PN MO9913720-A1.

```







```
XX (OHIS ) UNIV OHIO STATE.
PA
XX Ohashi N, Rikihisa Y;
PI
XX WPI; 1999-254290/21.
DR
XX P-PSDB; AAY06969.
DR
XX
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
XX
XX Discloure; Fig 29A; 55pp; English.
XX
XX The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
SQ Sequence 488 BP; 168 A; 75 C; 90 G; 155 T; 0 other;

Alignment Scores:
Pred. No.: 3.8e-45 Length: 488
Score: 500.50 Matches: 99
Percent Similarity: 71.26% Conservative: 20
Best Local Similarity: 59.28% Mismatches: 41
Query Match: 33.46% Indels: 7
DB: Gaps: 2

US-10-062-624-40 (1-293) x AAX34769 (1-488)
OY 127 SerTyrGluGluPheAspValLysAsnProGlyTyrThrLeuSerAspAlaTyrArg 146
DB 6 TCTTATGAAGATTGACGTAAATCTGGAAGAGTCTACT--ACAGACTCCATATGA 62
OY 147 TyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGluValSer 166
DB 63 TATTTCCGCTTGACAGTCGATGATGCTATATATCTCCACAGTCAGAAA----- 116
OY 167 AsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleLysValIleVal 186
DB 117 -----TTACTGTAATGAGAAACGACGGGTATGATCTCATCTGTATGATA 164
OY 187 AsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGly 206
DB 165 AATGGCTGTACAAATGCTCACTAAATGATATACAGACGAACTTACATATGTCACAGA 224
OY 207 AlaGlyValAspAlaIleGluPhePheAspValLysHisIleLysPheAlaTyrGlnSer 226
DB 225 CTAGAGAGAGATTATATAGATTCTTCATAGCTTCCTTACGTACGTACGTATCAGGT 284
OY 227 LysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyr 246
DB 285 AAGATGAGCATTAATCAATATATCCCTGAAGATGAGATTTATTTATGATGATATAC 344
OY 247 HisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGlnHisValAlaGluLeuAla 266
DB 345 CATTAAGTAAAGGCAAGTTTAAATTTTACAGTTCAACAGTACAGTACAGTACAGT 404
OY 267 SerIleProLysIleThrSerAlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGlu 286
DB 405 GCACCTCCCTAAAGTTATCATCTGACAGTTCGACACCTTAATATGATGATCTGTGTGA 464
OY 287 IleGlyAlaArgLeuThrPhe 293
DB 465 GCTGGAGTAAAGATTCAATATT 485

RESULT 15
AAX34744
ID AAX34744 standard; DNA; 852 BP.
XX
AC AAX34744;
```

```
XX
XX 05-JUL-1999 (first entry)
XX
XX DNA encoding OMP-1B protein.
DE
XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog; ss.
XX
XX OS Ehrlichia chaffeensis.
XX
XX W09913720-A1.
EN
XX
XX 25-MAR-1999.
PD
XX
XX 18-SEP-1998; 98WO-US19600.
PE
XX 19-SEP-1997; 97US-0059353.
FR
XX
XX (OHIS ) UNIV OHIO STATE.
XX
XX Ohashi N, Rikihisa Y;
PI
XX WPI; 1999-254290/21.
DR
XX P-PSDB; AAY06944.
DR
XX
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
XX
XX Discloure; Fig 4A; 55pp; English.
XX
XX The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
SQ Sequence 852 BP; 301 A; 155 C; 157 G; 239 T; 0 other;

Alignment Scores:
Pred. No.: 1.52e-44 Length: 852
Score: 498.00 Matches: 112
Percent Similarity: 55.48% Conservative: 60
Best Local Similarity: 36.13% Mismatches: 94
Query Match: 33.29% Indels: 44
DB: Gaps: 11

US-10-062-624-40 (1-293) x AAX34744 (1-852)
OY 1 MetAsnAsnLysLeuLysPheThrIleAsnThrValLeuValCysLeuLeuSerLeu 20
DB 1 ATGATTTCAAGAAATTTT-----CTAGCAGTGATTAATTTCTATATATCATC 54
OY 21 -----ProAsnIleSerSerLysValIleAsnAsnAsn 32
DB 55 TTACCTTACCAATCTTTTGACAGATCTGTAACTTCAATGATATACAGATCAACGACGC 114
OY 33 AlaLysLysTyrTyrGlyLeuTyrIleSerGlyGlnTyrLysProSerValSerValPhe 52
DB 115 AGAGAA-----GGCTTCTACATTAGGTAAAGTAAATCCAGCATATACACACTTC 165
OY 53 SerAsnProSerValLysGluThrAsnVal-----IleThrLysAsnLeu 67
DB 166 AGAAATTTCTAGCTGAAGAGCTCCCATCATGGAATACTTCTATCATCTAAAGATT 225
OY 68 IleAlaLeuLysLysAspValAspSerIleGluThrLysThrAspAlaSerValGlyIle 87
DB 226 TTCGGGCTGAAGAAAGACGAGAT-----ATA 252
OY 88 SerAsnProSerAsnPheThrIleProTyrThrAlaVal-----PheGlnAspAsnSerVal 106
DB 253 GCACATCTCGGAATTTTAAACAGACACATCCAGCCCTCGAGTTTCAGAAATTAACCTAATA 312
```



```

QY 107 Aen---PheANGlyThrIleGlyTyrThrPheAlaGluGlyThrArgValGluIleGlu 125
DB 313 TCAGATTCTCAGAAAGTATGTTATGCTGANG---GATGGCCAAAGATAAATTGA 369
QY 126 GlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThrLeuSer-----Asp 143
DB 370 GCTGATACCAAAATTTGATGCAAAATACTTGACAAACATGACATTAATGCGGTGAC 429
QY 144 AlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGlu 163
DB 430 TACTTAAATACTTGGACTATCTGCTGAA-----GACGCAATAGCAATAGAA 480
QY 164 LysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleLeuSer 183
DB 481 -----TATGTTGCTTAAATAAGAGGATCATCTTTATGCA 519
QY 184 ValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSerProTyrIle 203
DB 530 TTAATGGTTAACACTGCTATGATGACATTCACGCTGAGAGGATCTTTCATACCGTATGCA 579
QY 204 CysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIleLysPheAla 223
DB 580 TGTGCAGGTGTGAGACACACTTATTAACGTAATTAAGATTAAATTAATAATTCACA 639
QY 224 TyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSer 243
DB 640 TACCAAGGAAATAAGTATTAGCTATTCATCACACCAAGATTCCGCTTTATTTGA 699
QY 244 LeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGlnHisValAla 263
DB 700 GGATCTACACGAGGATTATAGAAATATTTAAACAAATACCTGTATATACACCTGTA 759
QY 264 GluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIleGlyTyrPhe 283
DB 760 GTATTAGAAAGAGCTCTTCAACCAATCTGCGCTAGTACTATTGACATCGGATATCTT 819
QY 284 GlyGlyGluIleGlyAlaArgLeuThrPhe 293
DB 820 GCGGAGAAAGTTGAGTAAAGTTCACTTC 849

```

RESULT 16  
ABK68876  
ID ABK68876 standard; DNA; 840 BP.  
AC ABK68876;  
DT 02-JUL-2002 (first entry)  
XX DNA encoding Ehrlichia canis p28-2.  
DE Ehrlichia canis infection; vaccine; serodiagnostic; gene; p28; ss;  
KW antibacterial.  
OS Ehrlichia canis.  
XX  
XX WO200222782-A2.  
XX  
XX 21-MAR-2002.  
XX  
XX 12-SEP-2001; 2001WO-US28759.  
XX  
XX 12-SEP-2000; 2000US-0660587.  
XX  
XX (RERE-) RES DEV FOUND.  
XX  
XX Walker DH, Yu X, McBride JW;  
XX  
XX WPI; 2002-351882/38.  
XX  
XX P-PsDB; AAU96116.  
XX  
XX New recombinant homologous 28 kilodalton immunodominant protein from  
XX  
XX Ehrlichia canis, useful for treating Ehrlichia canis infections -

```

PS Claim 5; Figure 14; 106bp; English.
XX The invention relates to a recombinant homologous 28 kDa immunodominant
CC protein, p28 (I) of Ehrlichia canis. (I), a 28-kDa antigen preferably
CC dispersed in a pharmaceutically acceptable carrier, is useful for
CC inhibiting E. canis infection in a subject. (I) is useful in the
CC development of vaccines and serodiagnostic tests that are particularly
CC effective for disease prevention and serodiagnosis. ABK6882-ABK68878
CC represent the 28-kDa antigen coding sequences and PCR primers of the
CC invention.
SQ Sequence 840 BP; 301 A; 141 C; 144 G; 254 T; 0 other;
XX
Alignment Scores:
Pred. No.: 9.74e-44 Length: 840
Score: 490.50 Matches: 114
Percent Similarity: 56.35% Conservative: 59
Best Local Similarity: 37.13% Mismatches: 85
Query Match: 32.79% Indels: 49
DB: Gaps: 12
US-10-062-624-40 (1-293) x ABK68876 (1-840)
QY 6 LysPheThrIleIleAsnThrValLeuValCysLeuLeuSer---LeuProAsnIleSer 24
DB 10 AAGAAATTCATGTAAGAGGCGCTTATCTCATTAATGTCATCACTTACATATCAGTCT 69
QY 25 SerSerLysAlaIle-----AsnAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40
DB 70 TTTCAGATCTCTGATGTTCAAGAACCTAATATATACAAAGAA-----GGCTTCTAC 120
QY 41 IleSerGlyIleTyrLysProSerValSerValPheSerAsnPheSerValLysGlu--- 59
DB 121 ATTATGCAAGTCAATCCAAATGATATACACTTTACAAATATCTCTGCAAGAACT 180
QY 60 -----ThrAsnValIleThrLysAsnLeuIleAlaLeuLysAspValAsp 75
DB 181 CCTATTAATGACAAATATCTCTCACTAAATAAGTTTCGACCTTAAGAAAGATGCTAT 240
QY 76 SerIleGluThrLysThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIle 95
DB 241 -----ATAACAAATAAAGACGATTTTACAGA 267
QY 96 ProTyrThrAlaVal---PheGlnAspAsnSerValAsn---PheANGlyThrIleGly 113
DB 268 GTAGCTCCAGCATTTGATTTCAAAATTAATTAATCAAGATTTTCAAGAAATTTGCT 327
QY 114 TyrThrPheAlaGluGlyThrArgValGluIleGlySerTyrGluGluPheAspVal 133
DB 328 TACTCTATNG---GACGACCAAGAAATAGACTTGAAGCTGATATCAACATTTAATCCA 384
QY 134 LysAsnPro-----GlyGlyTyrThrLeuSerAspAlaTyrArg 146
DB 385 AAAAACACCGATTAACAAATGATGATGATTAATGTGAATAC-----TATATAA 429
QY 147 TyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGluLysValSer 166
DB 430 CATTTGCAATATCTCGT-----AAAGATGCAATGAA 462
QY 167 AsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSerValIleVal 186
DB 463 GATCAGCAATAT---GTAATCTTAAATAATGACGCAATACTTTATATGCTATTAATGTT 519
QY 187 AsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGly 206
DB 520 AATACTTGCTATGACATTAACGTAAGAGGATATCTTCCATCATATGATGACAGGT 579
QY 207 AlaGlyValAspAlaIleGluPhePheAspValLeuHisIleLysPheAlaTyrGlnSer 226
DB 580 ATGAGAGAGATCTTATCACTATTTTAAAGACCTCAATCTAAATAATGCTTACCAAGGA 639
QY 227 LysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyr 246

```

Db 640 AAAATAGTATTAGTACCCCTATCAACACGAAAGTCTCTGATTTATGTTGATGATAC 699  
 Qy 247 HsHsValmeGlyAsnGlnPheLysAsnLeuValGlnHisValAlaGluLeuAla 266  
 Db 700 CATGCGTATTATGTAATAATTGAGAGATACCTGTAATACCTCTGATGATTTAAT 759  
 Qy 267 SerLeuProLysIleThrSerAlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGlu 286  
 Db 760 GATGCTCTCCAAACACATCTCTCTCAGTACTTTCAGCTTGACGTGGATCTTGGCGAGA 819  
 Qy 287 IlegIValArgLeuThrPhe 293  
 Db 820 ATGGAGATGAGGTTCACCTTC 840  
 RESULT 17  
 AAX34748  
 ID AAX34748 standard; DNA; 843 BP.  
 AC AAX34748;  
 XX  
 DT 05-JUL-1999 (first entry)  
 XX  
 DE DNA encoding OMP-1F protein.  
 XX  
 KM Outer membrane protein; OMP; *Escherichia chaffeensis*; *E. canis*; P30;  
 XX detection; dog; ss.  
 OS *Escherichia chaffeensis*.  
 XX  
 PN WO9913720-A1.  
 XX  
 PD 25-MAR-1999.  
 XX  
 PF 18-SEP-1998; 98WO-US19600.  
 XX  
 PR 19-SEP-1997; 97US-0059353.  
 XX  
 PA (OHIS ) UNIV OHIO STATE.  
 XX  
 PI Ohashi N, Rikihisa Y;  
 XX  
 DR WPI; 1999-254290/21.  
 XX  
 DR P-PSDB; AAY06948.  
 XX  
 PT Novel outer membrane proteins from *Escherichia chaffeensis* and  
 PT *Escherichia canis*  
 PS Disclosure; Fig 8A; 55pp; English.  
 XX  
 CC The invention provides isolated outer membrane proteins (OMP) from  
 CC *Escherichia chaffeensis* and *E. canis*. The *E. chaffeensis* proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1B to Z) shown.  
 CC in AAY06948-958. The *E. canis* proteins form part of the P30 family and  
 CC consist of proteins shown in AAY06959-970. The proteins and genes are  
 CC used to detect *E. chaffeensis* in patients and *E. canis* in dogs.  
 XX  
 SO Sequence 843 BP; 283 A; 128 C; 159 G; 273 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1.45e-41 Length: 843  
 Score: 470.50 Matches: 115  
 Percent Similarity: 56.034 Conservative: 57  
 Best Local Similarity: 37.464 Mismatches: 94  
 Query Match: 31.458 Indels: 41  
 DB: 20 Gaps: 14  
 US-10-062-624-40 (1-293) x AAX34748 (1-843)  
 Qy 1 McAAAsAsLysLeuLysPheThrIleIleAsnThrValLeuValCysLeuLeuSer--- 19  
 Db 1 ATGAATTGCAAA--AAATTT--TTATTAACAACATCAATAGTATCGGTAAATGCTTTC 54  
 Qy 20 LeuProAsnIleSerSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeu 39

Db 55 TTACTGGAAATATCATTTTTCATGATGACGATCAGAAACGAC--AATGTTGGTGAATTTTC 111  
 Qy 40 TyrIleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValGlyGlu 59  
 Db 112 TATATCAGTGGGAAATATGTAACAGTGTTCACATTTTGGCGTATCTCTGCTTAACAG 171  
 Qy 60 ThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThr 79  
 Db 172 GAAGAAATATACACACACCGAGTATTTGGATTTAAGCAAGTTGGATGGC----- 222  
 Qy 80 LysThrAspAlaSerValGlyLysSerAsnProSerAsn---PheThrIlePro---Tyr 97  
 Db 223 -----AGCACAAATATCTAAAAATTCCTCCAGAAATATACATTAAAGTTCCAAATAT 273  
 Qy 98 ThrAlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrThrPhe 116  
 Db 274 TCATTTAAATATATGAAATATATTCATTTCTAGTTCGAGAGCTGTGGTTAT---TTA 330  
 Qy 117 AlaGluGlyThrArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnPro 136  
 Db 331 ATGAATGCTCCAGAAATATGAGTTGAAATGCTCTATGAAATGAAATTTGATGTAACCAAG 390  
 Qy 137 GlyGlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGly 156  
 Db 391 GGTATATACATTAAGAAGATGCTCACAAATATATATGCTTTAACCCATTAACAGTGGGGA 450  
 Qy 157 AsnSerPheThrProLysGlyLysValSerAsnSer-----IlePheHisThr 172  
 Db 451 -----AAGCTAAGCAATGACAGTGTAAAGTTGTTT----- 483  
 Qy 173 ValMetArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysTyrAspPhe 192  
 Db 484 ---CTAAAAATGAAAGATCTCTGATATATCATCTATGTAATGATGATGATGATGATA 540  
 Qy 193 SerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIle 212  
 Db 541 ATAAGTGAAGAAATACCTTCTCTCTTACATATGTCAGAGTGTGTAAGATTTAATA 600  
 Qy 213 GluPhePheAspValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyr 232  
 Db 601 TCCATGTTTGAAGCTATTAACCTTAATATTTCTATACAGGAAGTATGAGTTAGGATAC 660  
 Qy 213 SerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLysValMeGlyAsn 252  
 Db 661 TCCATTAAGCCCAAGAGCTTCTTTTGTGTGTCGACATTTTCATTAAGTATAGCGAAT 720  
 Qy 253 GlnPheLysAsnLeuAsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThr 272  
 Db 721 GAATTCAGAGATATTCCT-----GCTATGATATCCAGTACCTCA 759  
 Qy 273 Ser-----AlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGlu 286  
 Db 760 ACTTCACAGGTAATCACTTACTATGTATACCTTAAGTATGATCCACTTGGAGTGGAA 819  
 Qy 287 IlegIValArgLeuThrPhe 293  
 Db 820 CTGGAGAGAGGTTTACTTT 840  
 RESULT 18  
 AAX34759  
 ID AAX34759 standard; DNA; 867 BP.  
 AC AAX34759;  
 XX  
 DT 05-JUL-1999 (first entry)  
 XX  
 DE DNA encoding P30 protein.  
 XX  
 KM Outer membrane protein; OMP; *Escherichia chaffeensis*; *E. canis*; P30;  
 XX detection; dog; ss.  
 OS *Escherichia canis*.

XX WO9913720-A1.  
 XX PD 25-MAR-1999.  
 XX PF 18-SEP-1998; 98WO-US19600.  
 XX PR 19-SEP-1997; 97US-0059353.  
 XX PA (OHIS ) UNIV OHIO STATE.  
 XX Ohashi N, Rikihisa Y;  
 XX WPI; 1999-254290/21.  
 XX DR P-PSDB; AAY06959.  
 XX Novel outer membrane proteins from Ehrlichia chaffeensis and  
 XX Ehrlichia canis  
 XX Disclosure; Fig 19A; 55pp; English.  
 XX The invention provides isolated outer membrane proteins (OMP) from  
 XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 XX of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 XX in AAY06943-958. The E. canis proteins form part of the P30 family and  
 XX consist of proteins shown in AAY06959-970. The proteins and genes are  
 XX used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX SQ Sequence 867 BP; 305 A; 150 C; 157 G; 255 T; 0 other;  
 XX

Alignment Scores:  
 Pred. No.: 1.51e-41 Length: 867  
 Score: 470.50 Matches: 120  
 Percent Similarity: 54.63% Conservative: 51  
 Best Local Similarity: 38.34% Mismatches: 97  
 Query Match: 31.45% Indels: 45  
 DB: 20 Gaps: 12

US-10-062-624-40 (1-293) x AAX34759 (1-867)

Oy 1 MetAsnLysLeuLysPheThrIleAsnThrValLeuValCysLeuLeuSer--- 19  
 Db 1 ATGATTCGCAAGATTTTC-----ATGCAAGTCGATTGATCATCTAATGCTTTC 54  
 Oy 20 LeuProAsnIleSerSerSerLysAla-----IleAsnAsnAlaLys 34  
 Db 55 TTACTAGCGTATCTTTTCTGATCAATACATGAGATATATATAATGGTAAC----- 108  
 Oy 35 LysTyrTyrGlyLeuTyrIleSerGlyGlnTyrLysProSerValSerValPheSerAsn 54  
 Db 109 -----TTTTACATTAGTGCAGAAATATATGCAAGTCGCTCACACTTTGGCGTA 156  
 Oy 55 PheSerValLysGluThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspVal 74  
 Db 157 TTTTCAGTTAAG 216  
 Oy 75 AspSerIleGluThrLysThrAspAlaSerValGlyIleSerAsnProSerAsn---Phe 93  
 Db 217 GACGAGCAACAATAAGAGATGCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTTC 276  
 Oy 94 ThrIlePro---TyrThrAlaValPheGlnAspLysSer---ValAsnPheAsnGlyThr 111  
 Db 277 TCCATTTCAAAATTTATCATTTTAAATATGAAACCAATCCATTTTGTAGGTTTGCAGGAGCT 336  
 Oy 112 IleGlyTyrThrPheAlaGluGlyThrArgValGluIleGluGlySerTyrGluLysPhe 131  
 Db 337 ATTGCTACTCAATGGGT---GGTCCAGGGGTAGAGTTTGAAGTGTCTTACGAATATTT 393  
 Oy 132 AspValLysAsnProGlyGlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAla 151  
 Db 394 GATGTAAACCAAGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 453  
 Oy 152 ArgGluMetLysGlyAsnSerPheThrProLysGluLysValSerAsnSerIlePheHis 171

Db 454 AGACACACCGGAGGT-----ATGCCACAACCGCGTCAATAAATTT----- 498  
 Oy 172 ThrValMetArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysTyrAsp 191  
 Db 499 GTCTTCTTAAATAATGAAGATTACTTACATATATGATGATTAAGCAAGCATGATGAT 558  
 Oy 192 PheSerLeuAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAla 211  
 Db 559 ATAACAATCGACAGCATGCCATTTTCCATATATATGTCAGGTATTGGTAGTACTTA 618  
 Oy 212 IleGluPhePheAspValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAla 231  
 Db 619 GTTTCGATGTTTGAACACTACAAATCTTAAATTTCTTATCAAGGAAATATTAGGTAGT 678  
 Oy 232 TyrSerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGly 251  
 Db 679 TACTCCATAGCCAGCAGCATCTGTTTTGTTGGAGGACACTTTTCAGAGTTATAGGT 738  
 Oy 252 AsnGlnPheLysAsnLeuAsnValGlnHisValAlaGluLeuAlaSerIleProLysIle 271  
 Db 739 AATGAATTTAAAGAC-----ATTCTTGCAATA 765  
 Oy 272 ThrSerAlaValAla-----ThrLeuAsnIle 280  
 Db 766 ACTCTCTGTCGAGCACAGAGAAATTAAGGCGACACAGTTTACACAGTAACATTAAACATA 825  
 Oy 281 GlyTyrPheGlyGlyGluIleGlyValArgLeuThrPhe 293  
 Db 826 TGCACCTTCGACTAGAGCTTGGAGCGAGTTTACTTTT 864

RESULT 19  
 AAC68705  
 ID AAC68705 standard; DNA; 843 BP.  
 XX AAC68705;  
 AC AAC68705;  
 XX  
 DT 02-MAR-2001 (first entry)  
 XX  
 DB Ehrlichia chaffeensis VSA4 gene partial coding sequence.  
 XX  
 KW Ehrlichia chaffeensis; VSA4; variable surface antigen 4; MAP1;  
 KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;  
 KW Rickettsia; Cowdria; Anaplasma; map2; lhwoif3; 4hwoif1; 18hwoif1;  
 KW 3gdorf3; ds.  
 XX  
 OS Ehrlichia chaffeensis.  
 XX  
 PN WO200065063-A2.  
 XX  
 XX 02-NOV-2000.  
 XX  
 PF 21-APR-2000; 2000WO-US10886.  
 XX  
 PR 22-APR-1999; 99US-0130725.  
 XX  
 PA (UVFL ) UNIV FLORIDA.  
 XX  
 PI Barbet AF, Bowie MV, Ganta RR, Burrridge MJ, Mahan SM, McGuire TC;  
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;  
 XX  
 XX WPI; 2000-679675/66.  
 DR P-PSDB; AAB36188.  
 XX  
 PT New polynucleotides useful as DNA vaccines for conferring immunity to  
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,  
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens  
 XX  
 PS Claim 4; Page 41; 63pp; English.  
 XX  
 CC The present sequence shows a high degree of similarity to the major  
 CC antigenic protein 1 (MAP1) gene of Ehrlichia sp. The MAP1 gene may be

CC used in a vaccines to protect animals or humans against rickettsial  
 CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,  
 CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response  
 CC protective against the rickettsial pathogen. The nucleic acid vaccines  
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.  
 CC Cowdria ruminatum genes designated map 2, lhwr3, 4hwr3, 18hwr3  
 CC and 3gdr3 may be used in therapeutic and diagnostic applications. The  
 CC polypeptides are useful for detecting antibodies associated with  
 CC infection by a rickettsial pathogen whilst the polynucleotides may be  
 CC used to detect the presence of rickettsial nucleic acids.

XX SQ Sequence 843 BP; 282 A; 127 C; 159 G; 275 T; 0 other;

## Alignment Scores:

Pred. No.: 2.4e-41 Length: 843  
 Score: 468.50 Matches: 115  
 Percent Similarity: 56.03% Conservative: 57  
 Best Local Similarity: 37.46% Mismatches: 94  
 Query Match: 31.32% Indels: 41  
 DB: 21 Gaps: 14

US-10-062-624-40 (1-293) x AAC68705 (1-843)

QY 1 MetAsnAsnLysLeuLysPheThrIleAenThrValLeuValCysLeuSer--- 19  
 DB 3 ATGAATTGCAAA---AAATTT---TTTAAACAACCTACATTAGTATCGCTAATGTCCTTC 56  
 QY 20 LeuProAsnIleSerSerLysAlaIleAsnAsnAsnAlaLysLysTyrTyrGlyLeu 39  
 DB 57 TTACCTGGAATATCATTTCTCGATGAGTACAGTACAGACGAC---AATGTGGTGTAATTC 113  
 QY 40 TyrIleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGlu 59  
 DB 114 TATATCAGTGGGAATATGTACCAAGTGTTCACATTTTGGCGGTATCTCTGCTAAACAG 173  
 QY 60 ThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThr 79  
 DB 174 GAAAGAAATACAAATCGGATGATTTGGATTAAAGCAAGATGGATGCGC----- 224  
 QY 80 LysThrAspAlaSerValGlyIleSerAsnProSerAsn---PheThrIlePro---Tyr 97  
 DB 225 -----AGCACAATATCTAAATAATCTCCAGAAATACATTTAAACGTTCCAAATAT 275  
 QY 98 ThrAlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrThrPhe 116  
 DB 276 TCATTTAAATATGAAATAATATCCATTTCTAGTGTTCAGGAGCTGTGGTTAT---TTA 332  
 QY 117 AlaGluGlyThrArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnPro 136  
 DB 333 ATGAATGTCGCAAGATAGATTAGNATGCTCTATGAACAACTTGTGTGAAACACG 392  
 QY 137 GlyGlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgLumMetLysGly 156  
 DB 393 GGTAAATAACTATAAGAACCATGCTCACAATATATTATGCTTTAACCCATAACAGTGGGA 452  
 QY 157 AsnSerPheThrProLysGluLysValSerAsnSer-----IlePheHisThr 172  
 DB 453 -----AAGCTAAGCAATGCAAGTGCATAGTTTGTGTTT----- 485  
 QY 173 ValMetArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysTyrAspPhe 192  
 DB 486 ---CTAAAAAATGAAGACTACTGTATATATACATTATGTTGAATGCATGCTATGTA 542  
 QY 193 SerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIle 212  
 DB 543 ATAAGTGAAGGAATACCTTCTCTCTACATATGTCAGGTCTTGTGCTGATTAAATA 602  
 QY 213 GluPhePheAspValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyr 232  
 DB 603 TCCATGTTTGAAGCTATAAACCCCTAAATTTCTTATCAAGGAAAGTAGTTTGAGTTAC 662  
 QY 233 SerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsn 252

DB 663 TCCATAAGCCAGAGCTTCTGTTTTTGGTGGACATTTTTCATAAGGTGATAGGGAAT 722  
 QY 253 GlnPheLysAsnLeuAsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThr 272  
 DB 723 GAATTCAGATATATCTCT-----GCTATGATACCCAGTACCTCA 761  
 QY 273 Ser-----AlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGlu 286  
 DB 762 ACTCTCAGGTAATCACTTTTACTATATAGTAACATAAGTGTATGCCACTTTGGAGTGAA 821  
 QY 287 IleGlyAlaArgLeuThrPhe 293  
 DB 822 CTTCGAGGAGGTTTAACTTT 842

RESULT 20  
 AAV07179  
 ID AAV07179 standard; DNA; 4683 BP.  
 XX  
 AC AAV07179;  
 XX  
 DT 14-SEP-1998 (first entry)  
 DE Ehrlichia chaffeensis VSA genomic locus.  
 KW MAP1 homologue; variable surface antigen; VSA1; VSA2; VSA3; VSA4;  
 KW VSA5; rickettsia; DNA vaccine; ss.  
 XX Ehrlichia chaffeensis.  
 PH Key Location/Qualifiers  
 CDS 3..134  
 FT /\*tag= a  
 FT /note= "VSA1 partial gene"  
 FT terminator 203..212  
 FT /\*tag= b  
 FT terminator 226..239  
 FT /\*tag= c  
 FT -35\_signal 349..354  
 FT /\*tag= d  
 FT -10\_signal 375..380  
 FT /\*tag= e  
 FT RBS 430..434  
 FT /\*tag= f  
 FT CDS 439..1299  
 FT /\*tag= g  
 FT /note= "VSA2"  
 FT terminator 1300..1309  
 FT /\*tag= h  
 FT terminator 1324..1333  
 FT /\*tag= i  
 FT misc\_feature 1347..1361  
 FT /\*tag= j  
 FT /note= "G-rich region"  
 FT -35\_signal 1473..1478  
 FT /\*tag= k  
 FT -10\_signal 1499..1504  
 FT /\*tag= l  
 FT RBS 1554..1558  
 FT /\*tag= m  
 FT CDS 1563..2399  
 FT /\*tag= n  
 FT /note= "VSA3"  
 FT terminator 2444..2456  
 FT /\*tag= o  
 FT terminator 2465..2477  
 FT /\*tag= p  
 FT misc\_feature 2491..2505  
 FT /\*tag= q  
 FT /note= "G-rich region"  
 FT -35\_signal 2618..2623  
 FT /\*tag= r  
 FT -10\_signal 2644..2649  
 FT /\*tag= s

```

FT RBS      2699..2703
FT /*tag= t
FT CDS      2708..3550
FT /*tag= u
FT /*note= "VSA4"
FT terminator 3586..3596
FT /*tag= y
FT terminator 3610..3622
FT /*tag= w
FT misc_feature 3586..3596
FT /*tag= x
FT /*note= "G-rich region"
FT -35_signal 3767..3772
FT /*tag= y
FT -10_signal 3788..3793
FT /*tag= z
FT RBS      3847..3851
FT /*tag= aa
FT CDS      3856..4683
FT /*tag= ab
FT /*note= "truncated VSA5 gene"
FT WO9816554-A1.
FT
FT 23-APR-1998.
FT
FT 17-OCT-1997; 97WO-US19044.
FT
FT 17-OCT-1996; 96US-0733230.
FT (UYFL ) UNIV FLORIDA.
FT
FT Barbet AF, Burrridge MJ, Ganta RR, Mahan SM, McGuire TC;
FT Nyika A, Rurangirwa FR;
FT
FT WPI: 1998-251232/22.
FT P-PSDB; AAW51091-05.
FT
FT Composition containing nucleic acid encoding rickettsial antigen -
FT useful for, e.g. stimulating protective immune response in humans or
FT animals
FT
FT Claim 4; Fig 2A-B; 39pp; English.
FT
FT This is the DNA sequence of a 4.6 kb genomic locus of Ehrlichia
FT chaffeensis that was obtained using a PCR cloning strategy based on
FT identifying genes homologous to the major antigenic protein MAP1
FT (see AAW51088) of Cowdria ruminantium. It includes 5 very similar
FT but non-identical open reading frames (ORFs), of which ORF1 is a
FT partial gene and ORF5 is nearly complete but lacks 5-7 amino acid
FT codons and a termination codon (see AAW51091-95). Due to their
FT similarity to MAP1 surface antigen genes of C. ruminantium, the
FT E. chaffeensis ORFs are designated variable surface antigen (VSA)
FT genes 1-5. A claimed composition comprises a nucleic acid (see
FT AAW07176-82) encoding a polypeptide (see AAW51088-99) that elicits a
FT protective immune response against a rickettsial pathogen. The
FT nucleic acid is used, in human or veterinary medicine, in vaccines
FT to protect against Rickettsia, Ehrlichia, Anaplasma and Cowdria
FT species. The nucleic acids are also useful as probes to identify
FT related sequences, e.g. for identification of organisms and for
FT diagnosing infection. Use of nucleic acid vaccines avoids the
FT problem of protein purification associated with protein-based
FT vaccines. The nucleic acid does not replicate in the host but
FT remains episomal and capable of expressing polypeptide for at least
FT 19 mth.
FT
FT SQ Sequence 4683 BP; 1590 A; 753 C; 818 G; 1522 T; 0 other;

```

```

Alignment Scores:
Pred. No.:      2,52e-40      Length:      4683
Score:          468.50      Matches:      112
Percent Similarity: 55.23%      Conservative: 57
Best Local Similarity: 36.60%      Mismatches: 98

```

```

Query Match: 31.32% Indels: 39
DB: 19 Gaps: 12
US-10-062-624-40 (1-293) x AAV07179 (1-4683)
QY 2 AsnAsnLysLeuLysPheThrIleAsnThrValLeuValCysLeuLeuSer---Leu 20
DB |||||
DB 2705 AATATGAATTGCAAAAAATTTTATACAACTACATTAGTATGCTAATGCTCTCTTA 2764
QY 21 ProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40
DB |||||
DB 2765 CCTGGATATCATTTTCTGATGCAGTACAGAACAC---AATGTTGGTGGTAATTTCTAT 2821
QY 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60
DB |||||
DB 2822 ATCAGTGGAAATATGACCAAGTGTTCACATTTGGCGTATTCTCTCTAAACAGGAA 2881
QY 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
DB |||||
DB 2882 AGAAATACAAATCGGAGTATTTGGATTAAAGCAAGATTGGGATGGC----- 2929
QY 81 ThrAspAlaSerValGlyIleSerAsnProSerAsn---PheThrIlePro---TyrThr 98
DB |||||
DB 2930 -----AGCACATATCTAAAAATCTCCAGAAATATACATTTAAGTTCCAAATTAITCA 2983
QY 99 AlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrThrPheAla 117
DB |||||
DB 2984 TTAAATATGAAATAATCCATTTCTAGGTTTTCAGGAGCTGTTGGTTAT---TTAATG 3040
QY 118 GluGlyThrArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGly 137
DB |||||
DB 3041 AATGCTCAAGATAGAGTTAGAAATGCTCTATGAAACATTTTGATGTGAAACACCGGTT 3100
QY 138 GlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsn 157
DB |||||
DB 3101 AATAACTATAGAAGCATGCTCACAAATATATCTTTAACCCATAACAGTGGGGGA--- 3157
QY 158 SerPheThrProLysGluLysValSerAsnSer-----IlePheHisThrVal 173
DB |||||
DB 3158 -----AAGCTAAGCAATGCAGGTGATAGTTTGTGTTT----- 3190
QY 174 MetArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysTyrAspPheSer 193
DB |||||
DB 3191 CTAAAAAATGAAGACTACTTGTATATATCACTTATGTTGAATGCTATGCTATGATGTAATA 3250
QY 194 LeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGlu 213
DB |||||
DB 3251 AGTGAAGGAATACCTTCTCTCTTACATATGTCAGGTGTTGTTGTTATTAATATCC 3310
QY 214 PhePheAspValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSer 233
DB |||||
DB 3311 ATGTTTGAAGCTATAAACCCCTAAATTTCTTATCAAGGAAGTTAGTTTGAGTTACTCC 3370
QY 234 LeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGln 253
DB |||||
DB 3371 ATAAGCCAGGAAGCTTCTGTTTGTGTTGGTGGACATTTTCATAAGGTGATAGGGAATGAA 3430
QY 254 PheLysAsnLeuAsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSer 273
DB |||||
DB 3431 TTCAGATATATTCCT-----GCTATGATACCAGTACCTCAACT 3469
QY 274 -----AlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGluIle 287
DB |||||
DB 3470 CTCACAGGTAATCACTTTACTATAGTAACACTAGTGTATGCCACTTTGGAGTGGAACTT 3529
QY 288 GlyAlaArgLeuThrPhe 293
DB |||||
DB 3530 GGAGGAAGGTTTAACCTTT 3547

```

RESULT 21

AAC68716

ID AAC68716 standard; DNA; 4683 BP.

XX

AC AAC68716;  
 XX 02-WAR-2001 (first entry)  
 DT Ehrlichia chaffeensis 28 kDa gene locus.  
 DE  
 XX Ehrlichia chaffeensis; MAP1; major antigenic protein 1; antirickettsial;  
 KW vaccine; gene therapy; Rickettsia; Cowdria; Anaplasma; map2; ihworf3;  
 KW 4hworf1; 18hworf1; 3gdorf3; ds.  
 XX  
 OS Ehrlichia chaffeensis.  
 XX  
 XX WO2000065063-A2.  
 PN 02-NOV-2000.  
 XX  
 XX 21-APR-2000; 2000WO-US10886.  
 PF  
 XX 22-APR-1999; 99US-0130725.  
 PR  
 XX (UYFL) UNIV FLORIDA.  
 PA  
 XX Barbet AF, Bowie MV, Ganta RR, Burrige MJ, Mahan SM, McGuire TC;  
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmore WW, Alleman AR;  
 PI  
 XX WPI; 2000-679675/66.  
 DR  
 XX New polynucleotides useful as DNA vaccines for conferring immunity to  
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,  
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens  
 PT  
 XX  
 PS Example 2; Fig 2A-2B; 63pp; English.  
 CC  
 CC The present sequence is given in a specification relating to nucleic  
 CC acid vaccines which may be used to protect animals or humans against  
 CC rickettsial diseases caused by a organisms of Rickettsia sp.,  
 CC Ehrlichia sp., Anaplasma sp. or Cowdria sp. The vaccine elicits an  
 CC immune response protective against the rickettsial pathogen. The  
 CC vaccine comprises the major antigenic protein 1 (MAP1) or major  
 CC antigenic protein 2 (MAP2) of rickettsial pathogens. The nucleic acid  
 CC vaccines can be driven by the human cytomegalovirus (HCMV)  
 CC enhancer-promoter. Cowdria ruminantium genes designated map 2, ihworf3,  
 CC 4hworf1, 18hworf1 and 3gdorf3 may be used in therapeutic and diagnostic  
 CC applications. The polypeptides are useful for detecting antibodies  
 CC associated with infection by a rickettsial pathogen whilst the  
 CC polynucleotides may be used to detect the presence of rickettsial  
 CC nucleic acids.  
 XX  
 SQ Sequence 4683 BP; 1576 A; 747 C; 831 G; 1526 T; 3 other;  
 XX  
 Alignment Scores:  
 Pred. No.: 2,52e-40 Length: 4683  
 Score: 468.50 Matches: 112  
 Percent Similarity: 55.23% Conservative: 57  
 Best Local Similarity: 36.60% Mismatches: 98  
 Query Match: 31.32% Indels: 39  
 DB: 21 Gaps: 12  
 US-10-062-624-40 (1-293) x AAC68716 (1-4683)  
 QY 2 AsnAsnLysLeuLysPheThrIleAsnThrValLeuValCysLeuLeuSer---Leu 20  
 DB 2705 ATATGAAATGCAAAATTTTATACAACTACATTAAGTATGCTGCTCTCTTA 2764  
 QY 21 ProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40  
 DB 2765 CTGGATATATCTTCTGATGAGTACAGACGAC---AAGTTGGTGAATTTCTAT 2821  
 QY 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60  
 DB 2822 ATCAGTGGGAATATGTCACCAAGTGTTCACATTTTGGGTATCTCTGCTAAACAGGAA 2881

QY 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80  
 DB 2882 AGAAATCAACAATCGGAGTATTGGATTAAAGCAAGATTGGATGGC----- 2929  
 QY 81 ThrAspAlaSerValGlyIleSerAsnProSerAsn---PheThrIlePro---TyrThr 98  
 DB 2930 -----AGCAATATCTAAAAATTCCTCCAGAAAATACATTTAAGTTCCTCAAAATATTATCA 2983  
 QY 99 AlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrPheAla 117  
 DB 2984 TTAAATATGAAATAATATCCATTTCTAGGTTTTCAGAGAGCTGTGGTTAT---TTAATG 3040  
 QY 118 GluGlyThrArgValGluIleGlySerTyrGluGluPheAspValLysAsnProGly 137  
 DB 3041 AATGTCCTCAAGATAGAGTTAGAAATGCTCTATGAACATTTGATGTGAAAACACGAGGT 3100  
 QY 138 GlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsn 157  
 DB 3101 AATTAATATAAGAACGATGCTCACAAATATTATGCTTTAACCCATAACAGTGGGGA--- 3157  
 QY 158 SerPheThrProLysGluLysValSerAsnSer-----IlePheHisThrVal 173  
 DB 3158 -----AAGCTAAGCAATGACAGTGTGATAAGTTTGTGTTT----- 3190  
 QY 174 MetArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysTyrAspPheSer 193  
 DB 3191 CTAAAAATAGAGGACTACTTGATATATCATCTTAATGTAATGCTATGATGATGATAATA 3250  
 QY 194 LeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGlu 213  
 DB 3251 AGTGAAGAAATACCTTTCTCTCTTACATATGTGCAGGTGTGGTACTGATTAAATATCC 3310  
 QY 214 PhePheAspValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSer 233  
 DB 3311 ATGTTGAAGCTATATAAACCTTAAATTTCTTATCAAGGAAAGATTAGGTTTGAGTTACTCC 3370  
 QY 234 LeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGln 253  
 DB 3371 ATAAGCCAGAGAGCTTCTGTTTTTGTGGTGACATTTTTCATAGGTGATAGGGAATGAA 3430  
 QY 254 PheLysAsnLeuAsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSer 273  
 DB 3431 TTCAGAGATATTCTCT-----GCTATGATACCCAGTACCTCAACT 3469  
 QY 274 -----AlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGluLe 287  
 DB 3470 CTCACAGTAATACCTTTTACTATAGTAACACTTAAGTGTATGCCACTTTGGAGTGAACCTT 3529  
 QY 288 GlyAlaArgLeuThrPhe 293  
 DB 3530 GGAGGAAGGTTTAACTTT 3547  
 RESULT 22  
 AAS07578  
 ID AAS07578 standard; DNA; 4683 BP.  
 XX  
 AC AAS07578;  
 XX  
 DT 23-OCT-2001 (first entry)  
 DE DNA encoding variable surface antigens 1-5 (VSA1-5) from E. chaffeensis.  
 XX  
 KW Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;  
 KW infection; heartwater; diagnostic; variable surface antigen; VSA; ds.  
 XX  
 OS Ehrlichia chaffeensis.  
 XX  
 XX Location/Qualifiers  
 FT CDS 3..134  
 FT /tag= a  
 FT /product= "Variable surface antigen 1 (VSA1)"  
 FT /partial  
 FT /note= "No start codon"

```

FT terminator 179..188
FT /tag= b
FT /note= "Transcription terminator of VSA1 gene"
FT terminator 202..212
FT /tag= c
FT /note= "Transcription terminator of VSA1 gene"
FT GC_signal 226..239
FT /tag= d
FT /note= "GC-rich sequence of VSA2 gene"
FT -35_signal 349..354
FT /tag= e
FT /note= "-35 region of VSA2 gene"
FT -10_signal 375..380
FT /tag= f
FT /note= "-10 region of VSA2 gene"
FT RBS 430..434
FT /tag= g
FT /note= "Ribosome binding site of VSA2 gene"
FT CDS 439..1299
FT /tag= h
FT /product= "Variable surface antigen 2 (VSA2)"
FT terminator 1300..1309
FT /tag= i
FT /note= "Transcription termination signal of VSA2"
FT GC_signal 1324..1333
FT /tag= j
FT /note= "Transcription termination signal of VSA2"
FT -35_signal 1349..1361
FT /tag= k
FT /note= "GC-rich sequence of VSA3 gene"
FT -10_signal 1473..1478
FT /tag= l
FT /note= "-35 region of VSA3 gene"
FT RBS 1499..1504
FT /tag= m
FT /note= "-10 region of VSA3 gene"
FT CDS 1554..1558
FT /tag= n
FT /note= "Ribosome binding site of VSA3 gene"
FT terminator 1563..2399
FT /tag= o
FT /product= "variable surface antigen 3 (VSA3)"
FT terminator 2443..2455
FT /tag= p
FT /note= "Transcription termination signal of VSA3"
FT terminator 2465..2477
FT /tag= q
FT /note= "Transcription termination signal of VSA3"
FT GC_signal 2491..2505
FT /tag= r
FT /note= "GC-rich sequence of VSA4 gene"
FT -35_signal 2618..2623
FT /tag= s
FT /note= "-35 region of VSA4 gene"
FT -10_signal 2644..2649
FT /tag= t
FT /note= "-10 region of VSA4 gene"
FT RBS 2699..2703
FT /tag= u
FT /note= "Ribosome binding site of VSA4 gene"
FT CDS 2708..3550
FT /tag= v
FT /product= "Variable surface antigen 4 (VSA4)"
FT terminator 3586..3596
FT /tag= w
FT /note= "Transcription termination signal of VSA4"
FT terminator 3610..3622
FT /tag= x
FT /note= "Transcription termination signal of VSA4"
FT GC_signal 2491..2505
FT /tag= y
FT /note= "GC-rich sequence of VSA5 gene"
FT -35_signal 3767..3772
FT /tag= z
FT /note= "-35 region of VSA5 gene"
FT -10_signal 3788..3793
FT /tag= aa
FT /note= "-10 region of VSA5 gene"
FT RBS 3847..3851
FT /tag= ab
FT /note= "Ribosome binding site of VSA5 gene"
FT CDS 3856..4683
FT /tag= ac
FT /product= "Variable surface antigen 5 (VSA5)"
FT /partial
FT /note= "No stop codon"
XX
XX US6251872-B1.
PN
XX
XX 26-JUN-2001.
XX
XX 17-OCT-1997; 97US-09533326.
XX
XX 17-OCT-1996; 96US-0733230.
XX (UYFL ) UNIV FLORIDA.
XX
XX Barbet AF, Ganta RR, McGuire TC, Burrigge MJ, Nyika A;
XX Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
XX WPI; 2001-424487/45.
XX P-PSDB; AAU04195, AAU04196, AAU04197, AAU04198, AAU04199.
XX
XX New MAP2 genes and polypeptides useful as vaccines for conferring
XX immunity to human and animal rickettsial diseases, e.g. heartwater, or
XX as molecular markers in nucleic acid analysis procedures
XX
XX Example 3; Fig 2A-2B; 30pp; English.
XX
XX The sequence represents the coding sequence of variable surface antigen
XX (VSA) gene locus encoding VSA1-5 proteins of Ehrlichia chaffeensis, which
XX have similarity to major antigen proteins (MAP). The MAP polynucleotides
XX and polypeptides are useful as vaccines for conferring immunity to
XX rickettsia infection, including Cowdria ruminantium causing heartwater.
XX The MAP polynucleotides may be used as molecular markers in nucleic acid
XX analysis procedures, and to produce the MAP polypeptides, which may
XX be used to raise antibodies that are reactive with the polypeptides.
XX The nucleic acids may further be used as probes to identify
XX complementary sequences within other nucleic acid molecules or genomes,
XX where such probes can be applied to identify or distinguish infectious
XX strains of organisms in diagnostic procedures or in rickettsial
XX research where identification of particular organisms or strains is
XX needed.
XX
XX SQ Sequence 4683 BP; 1578 A; 739 C; 833 G; 1533 T; 0 other;

Alignment Scores:
Pred. NO.: 2.52e-40 Length: 4683
Score: 468.50 Matches: 112
Percent Similarity: 55.23% Conservative: 57
Best Local Similarity: 36.60% Mismatches: 98
Query Match: 31.32% Indels: 39
DB: 22 Gaps: 12

US-10-062-624-40 (1-293) x AAS07578 (1-4683)
Qy 2 AenAsnLysLeuLysPheThrIleAenThrValLeuValCysLeuLeuSer--Leu 20
Dbb 2705 AATATGAATTGCAGAAAATTTTATACAACTACATTAGTATCGCTAATGCTCTTCTTA 2764
Qy 21 ProAsnIleSerSerSerLysAlaIleAsnAsnAlaLysLysTyTyxGlyLeuTyx 40
Dbb 2765 CCTGGAATACATTTTCTGATGCAGTACAGACACAC--AATGTTGGTGAATTTCTAT 2821
Qy 41 IleSerGlyGlnTyLysProSerValSerValSerAenPheSerAenPheSerValGluThr 60

```

```

Db      2822 ATCAGTGGGAATATGTACCAAGTGTTCACATTTTGGCGTATTCTCTGCTTAAACAGGAA 2881
Qy      61  AenValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
Db      2882 AGAAATACCAACATCGGAGTATTGGATTAAAGCAAGATTGGATGGC-----2929
Qy      81  ThrAspAlaSerValGlyIleSerAsnProSerAsn---PheThrIlePro---TyrThr 98
Db      2930 -----AGCACAAATCTTAAATAATCTCCAGAAATATACATTTAAACGTTCCAAATATTATCA 2983
Qy      99  AlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrThrPheAla 117
Db      2984 TTTAAATATGAAATAATCCATTTCTAGGTTTTCAGGAGCTGTGGTTAT---TTAATG 3040
Qy      118 GluGlyThrArgValGluGlySerTyrGluPheAspValLysAsnProGly 137
Db      3041 AATGGTCCAAAGATAGATTAGAAATGCTCTATGAACATTTTGTGTGAAAAACAGGCT 3100
Qy      138 GlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsn 157
Db      3101 AATAACTATAAGACGATGCTCACAAATATTATGCTTTAAACCCATAACAGTGGGGA---3157
Qy      158 SerPheThrProLysGluLysValSerAsnSer-----IlePheHisThrVal 173
Db      3158 -----AAGCTAAGCAATGAGGTGATAAGTTGTTTT-----3190
Qy      174 MetArgAsnAspGlyLeuSerIleLeuSerValIleValAsnValCysTyrAspPheSer 193
Db      3191 CTAAAAATGAAGGACTACTTGATATATCACTATGTTGAATGCATGCTATGATGTAATA 3250
Qy      194 LeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGlu 213
Db      3251 AGTGAAGGAATACCTTCTCTCTACATATGTCAGGTGTTGGTACTGATTAAATATCC 3310
Qy      214 PhePheAspValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSer 233
Db      3311 ATGTTTGAAGCTATAAACCTTAATTTCTTATCAGGAAGTAGTTAGTTTGTACTCC 3370
Qy      234 LeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGln 253
Db      3371 ATAAGCCCAAGAGCTCTCTGTTTGTGGTGGACATTTTTCATAAGGTGATAGGAATGAA 3430
Qy      254 PheLysAsnLeuAsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSer 273
Db      3431 TTCAGAGATATTCTCT-----GCTATGATACCCAGTACCTCAACT 3469
Qy      274 -----AlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGluIle 287
Db      3470 CTCACAGGTAAATCACTTTACTATAGTAACACTAAGTGTATGCCACTTTGGAGTGAACCTT 3529
Qy      288 GlyAlaArgLeuThrPhe 293
Db      3530 GGAGGAAGGTTTAACTTT 3547

RESULT 23
AAV07176
ID      AAV07176 standard; DNA; 864 BP.
XX
AC      AAV07176;
XX
DT      14-SEP-1998 (first entry)
XX
DE      Cowdria ruminantium MAP1 gene coding sequence.
XX
KW      MAP1 gene; major antigenic protein 1; rickettsia; heartworm;
XX      DNA vaccine; ss.
XX
OS      Cowdria ruminantium.
XX
PN      WO9816554-A1.
XX
PD      23-APR-1998.
XX

```

```

PF      17-OCT-1997; 97WO-US19044.
XX
PR      17-OCT-1996; 96US-0733230.
XX
PA      (UYFL ) UNIV FLORIDA.
XX
PI      Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;
PI      Nyika A, Rurangirwa FR;
XX
XX      WPI; 1998-251232/22.
DR      P-PSDB; AAW51088.
XX
PT      Composition containing nucleic acid encoding rickettsial antigen -
PT      useful for, e.g. stimulating protective immune response in humans or
PT      animals
XX
XX      Claim 4; Page 14-15; 39pp; English.
XX
CC      This DNA molecule comprises the coding region of the major antigen
CC      protein 1 gene (MAP1) of Cowdria ruminantium, the causative agent
CC      of heartwater in domestic ruminants. It codes for a 287-amino acid
CC      MAP1 protein (see AAW51088). A claimed composition comprises a
CC      nucleic acid (see AAV07176-82) encoding a polypeptide (see AAW51088-99)
CC      that elicits a protective immune response against a rickettsial
CC      pathogen. The nucleic acid is used, in human or veterinary
CC      medicine, in vaccines to protect against Rickettsia, Ehrlichia,
CC      Anaplasma and Cowdria species. The nucleic acids are also useful
CC      as probes to identify related sequences, e.g. for identification of
CC      organisms and for diagnosing infection. Use of nucleic acid
CC      vaccines avoids the problem of protein purification associated with
CC      protein-based vaccines. The nucleic acid does not replicate in the
CC      host but remains episomal and capable of expressing polypeptide for
CC      at least 19 mch.
XX
SQ      Sequence 864 BP; 307 A; 139 C; 149 G; 269 T; 0 other;

Alignment Scores:
Pred. No.:      2,666-40      Length:      864
Score:          459.00      Matches:    110
Percent Similarity: 57.38%      Conservative: 65
Best Local Similarity: 36.07%      Mismatches: 100
Query Match:    30.68%      Indels:     30
DB:             19          Gaps:       10

US-10-062-624-40 (1-293) x AAV07176 (1-864)
Qy      1  MetAsnAsnLysLeuLysPheThrIleAsnThrValLeuValCysLeuSer--- 19
Db      1  ATGAATTGCAAGAAAAATTTT-----ATCAAGTACACTAATATCATTTAGTGTCAATTT 54
Qy      20  LeuProAsnIleSerSerSerLysLysAlaIleAsnAsnAsnAlaLysLysTyrTyrGlyLeu 39
Db      55  TTACCTGGTGTGTCCTTTTCTGATGTAATACAGGAAGACAGCAACCCAGCAGGAGTGT 114
Qy      40  TyrIleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGlu 59
Db      115  TACATTAGCGCAAAATACATGCCAACTGCATCACATTTTGGTAAATGTCAATCAAGAA 174
Qy      60  ThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThr 79
Db      175  GATTCAAAAAATACTCAACGCGTATTGCTCTAAAAAAGAGATTGGATGGCGTTAAACA 234
Qy      80  LysThrAspAlaSerValGlyIleSerAsnProSerAsnPheThr---IleProTyrThr 98
Db      235  CCATCAGATTCTAGC-----AATACTAATTTCTACAATTTTACTGAAAGAGACTATTCT 288
Qy      99  AlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrThrPheAla 117
Db      289  TTCAGATATGAACAAATCCCGTTTGTAGGTTTCGCTGGAGCAATTTGGGTACTCAATG--- 345
Qy      118  GluGlyThrArgValGluGlySerTyrGluPheAspValLysAsnProGly 137
Db      346  AATGACCAAGAAATAGAGTTTCGAAGTATCTATGAACATTTTGTATGAATTTTGTATGAACCTTAGGT 405

```



Qy 138 GlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaAArgGluMetLysGlyAsn 157  
 Db 406 GGCACTATAAAACACGACACATGACTGTGCTTTA-----GATACAGCAGCACA 459  
 Qy 158 SerPheThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAsp 177  
 Db 460 AATGACATATGCGCGAGGATTAACATCTGTT-----ATGTTAAACACGAA 510  
 Qy 178 GlyLeuSerIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeu 197  
 Db 511 AATTAAACAATATATCAATTAATGTTAAATGCGTGTATGATATCATCTGTTGATGGAATA 570  
 Qy 198 SerIleSerProTyrIleCysGlyAlaGlyValAspAlaIleGluPhePheAspVal 217  
 Db 571 CCAGTTTCTCCATATGATGTGCAGGTATGCGACTGACTGATGCTAGTCAATTAATGCT 630  
 Qy 218 LeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsn 237  
 Db 631 ACAATCTTAATTAATCTTATCAAGGAAGCTAGGCAATGTTACTCAATCAATTTCTGAA 690  
 Qy 238 IleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeu 257  
 Db 691 GCTTCTATCTTATCGGTGGACATTTCCATAGAGTTATAGGTATGATTAATTAATA----- 744  
 Qy 258 AsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSer----- 273  
 Db 745 -----GATATTGCTACTTAAATAATATTACTTCAAAACACAGGAATA 786  
 Qy 274 -----AlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGluIleGly 288  
 Db 787 TCTAATCTGCTTGGTCATCAGCACTTGTATGTTGCTCACTTGTGTTATGAAATTTGGA 846  
 Qy 289 AlaArgLeuThrPhe 293  
 Db 847 GGAAGTTTGTATTT 861  
 RESULT 24  
 AAC68699  
 ID AAC68699 standard; DNA; 864 BP.  
 XX AC AAC68699;  
 XX AC  
 DT 02-MAR-2001 (first entry)  
 XX DE  
 XX KW Cowdria ruminatum MAP1 gene.  
 KW Cowdria ruminatum; MAP1; major antigenic protein 1; antirickettsial;  
 KW vaccine; gene therapy; Rickettsia; Ehrlichia; Anaplasma; map2; ihworf3;  
 KW 4hworf1; 18hworf1; 3gdorf3; ds.  
 XX OS  
 XX OS Cowdria ruminatum.  
 XX FN W0200065063-A2.  
 XX PD 02-NOV-2000.  
 XX PF 21-APR-2000; 2000WO-US10886.  
 XX PR 22-APR-1999; 99US-0130725.  
 XX PA (UYFL) UNIV FLORIDA.  
 XX PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;  
 XX PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmore WW, Allemen AR;  
 XX DR WPI; 2000-679675/66.  
 XX DR P-PSDB; AAB36182.  
 PT New polynucleotides useful as DNA vaccines for conferring immunity to  
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminatum,  
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens

## XX Claim 4; Fig 1; 63pp; English.

XX The present sequence is given in a specification relating to nucleic  
 CC acid vaccines containing genes to protect animals or humans against  
 CC rickettsial diseases caused by a organisms of Rickettsia sp., Ehrlichia  
 CC sp., Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response  
 CC protective against the rickettsial pathogen. The vaccine comprises the  
 CC major antigenic protein 1 (MAP1) gene or the major antigenic protein 2  
 CC (MAP2) gene of rickettsial pathogens. The nucleic acid vaccines can be  
 CC driven by the human cytomegalovirus (HCMV) enhancer-promoter. Cowdria  
 CC ruminatum genes designated map 2, ihworf3, 4hworf1, 18hworf1 and  
 CC 3gdorf3 may be used in therapeutic and diagnostic applications. The  
 CC polypeptides are useful for detecting antibodies associated with  
 CC infection by a rickettsial pathogen whilst the polynucleotides may be  
 CC used to detect the presence of rickettsial nucleic acids.

SQ Sequence 864 BP; 307 A; 139 C; 149 G; 269 T; 0 other;

## Alignment Scores:

Pred. No.: 2,66e-40 Length: 864  
 Score: 459.00 Matches: 110  
 Percent Similarity: 57.38% Conservative: 65  
 Best Local Similarity: 36.07% Mismatches: 100  
 Query Match: 30.88% Indels: 30  
 DB: 21 Gaps: 10

US-10-062-624-40 (1-293) x AAC68699 (1-864)

Qy 1 MetAsnAsnLysLeuLysPheThrIleIleAsnThrValLeuValCysLeuLeuSer--- 19  
 Db 1 ATGAATTTGCAAGAAAATTTTT-----ATCACAAGTACACTAATATCATGTAGTGTCTTT 54  
 Qy 20 LeuProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeu 39  
 Db 55 TTACTGGTGTGCTCTTTTCGATGTAATACAGGAACAGCACCCAGCAGGAGTGT 114  
 Qy 40 TyrIleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGlu 59  
 Db 115 TACATTAGCGCAAAATACATGCGCAACTGCATCACAATTTTGGTAAATGTCAATCAAGAA 174  
 Qy 60 ThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThr 79  
 Db 175 GATTCAAAAAATACTCAACGGTATTTGGTCTAAAAAAGATTGGGATGGCGTTAAAAACA 234  
 Qy 80 LysThrAspAlaSerValGlyIleSerAsnProSerAsnPheThr---IleProTyrThr 98  
 Db 235 CCATCAGATCTAGC-----AATACATATTTCTACAATTTTCTGAAAAAGACTATCT 288  
 Qy 99 AlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrThrPheAla 117  
 Db 289 TTCAGATATGAACAAATCCGTTTTTAGGTTTCGCTGGAGCAATGGGTACTCAATG--- 345  
 Qy 118 GluGlyThrArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGly 137  
 Db 346 AATGGACCAAGATAGAGTTCGAAGTATCCTATGAAACTTTTGTATGTAATAAAGACCTAGGT 405  
 Qy 138 GlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaAArgGluMetLysGlyAsn 157  
 Db 406 GGCAACTATAAAACACGACACATCATGCTGCTTTA-----GATACAGCAGCACA 459  
 Qy 158 SerPheThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAsp 177  
 Db 460 AATGACATTAATGCGCGAGGATTAACATCTGTT-----ATGTTAAACACGAA 510  
 Qy 178 GlyLeuSerIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeu 197  
 Db 511 AATTAAACAATATATCAATTAATGTTAAATGCGGTGTTATGATATCATCTGTTGATGGAATA 570  
 Qy 198 SerIleSerProTyrIleCysGlyAlaGlyValAspAlaIleGluPhePheAspVal 217  
 Db 571 CCAGTTTCTCCATATGATGTGCAGGTATGCGCACTGACTGACTAGTGTAGTCAATTAATGCT 630

QY 218 LeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsn 237  
 Db 631 ACAATCTTAATTAATCTATACAGAAAGCTAGGCATTAAGTACTCAATCAATCTGAA 690  
 QY 238 IleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeu 257  
 Db 691 GCTTCTATCTTTATCGGTGGACATTTCCATAGAGTTATAGGTAAATGAATTTAAA----- 744  
 QY 258 AsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSer----- 273  
 Db 745 -----GATATTGCTACCTTTAAAAATATTACTTCAAAAACAGGAATA 786  
 QY 274 -----AlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGluIleGly 288  
 Db 787 TCTAATCTGGCTTTGGATCAGCAACACTTGAATGTTGTCACCTTTGGTATAGAAATTTGA 846  
 QY 289 AlaArgLeuThrPhe 293  
 Db 847 GGAAGTTGTGATTT 861  
 RESULT 25  
 ID AAS07575 standard; DNA; 864 BP.  
 XX  
 AC AAS07575;  
 XX  
 DT 23-OCT-2001 (first entry)  
 XX  
 DE DNA encoding major antigenic protein 1 (MAP1).  
 XX  
 KW Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;  
 KW infection; heartwater; diagnostic; ds.  
 XX  
 OS Cowdria ruminantium.  
 XX  
 FH Key  
 FT 1..864  
 CDS  
 FT /\*tag= a  
 FT /product= "Major antigenic protein 1 (MAP1)"  
 XX  
 US6251872-B1.  
 XX  
 PD 26-JUN-2001.  
 XX  
 PF 17-OCT-1997; 97US-0953326.  
 XX  
 PR 17-OCT-1996; 96US-0733230.  
 XX  
 PA (UYFL ) UNIV FLORIDA.  
 XX  
 PI Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;  
 PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;  
 XX  
 DR WPI; 2001-424487/45.  
 DR P-PSDB; AAU04192.  
 XX  
 PT New MAP2 genes and polypeptides useful as vaccines for conferring  
 PT immunity to human and animal rickettsial diseases, e.g. heartwater, or  
 PT as molecular markers in nucleic acid analysis procedures  
 XX  
 PS Example 1; Fig 1; 30pp; English.  
 XX  
 CC The sequence represents the coding sequence of major antigenic protein 1  
 CC (MAP1) isolated from Cowdria ruminantium. The MAP polynucleotides and  
 CC polypeptides are useful as vaccines for conferring immunity to rickettsia  
 CC infection, including Cowdria ruminantium causing heartwater. The MAP  
 CC polynucleotides may be used as molecular markers in nucleic acid  
 CC analysis procedures, and to produce the MAP polypeptides, which may  
 CC be used to raise antibodies that are reactive with the polypeptides.  
 CC The nucleic acids may further be used as probes to identify  
 CC complementary sequences within other nucleic acid molecules or genomes,  
 CC where such probes can be applied to identify or distinguish infectious  
 CC strains of organisms in diagnostic procedures or in rickettsial

CC research where identification of particular organisms or strains is  
 CC needed.  
 XX Sequence 864 BP; 307 A; 139 C; 149 G; 269 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 2.66e-40 Length: 864  
 Score: 459.00 Matches: 110  
 Percent Similarity: 57.38% Conservative: 65  
 Best Local Similarity: 36.07% Mismatches: 100  
 Query Match: 30.68% Indels: 30  
 Gaps: 10  
 DB: 22  
 US-10-062-624-40 (1-293) x AAS07575 (1-864)  
 QY 1 MetAsnLysLeuLysPheThrIleAsnThrValLeuValCysLeuLeuSer--- 19  
 Db 1 ATGAATTGCAAGAAAATTTT-----ATCAAGATACACTAATATCATTTAGTGTCTTT 54  
 QY 20 LeuProAsnIleSerSerLysAlaIleAsnAsnAsnAlaLysLysTyrGlyLeu 39  
 Db 55 TTACTGTGTGTCTCTTTTCTGATGTAATACAGGAAGACAGCAACCCAGCGAGGTGT 114  
 QY 40 TyrIleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGlu 59  
 Db 115 TACATTAGCGCAAAATACATCCCACTCCATCATCATTTTGGTAAATGTCAATCAAGAA 174  
 QY 60 ThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThr 79  
 Db 175 GATTCAAAAAATACTCAAAACGGTATTTGGTCTAAAAAAGATTGGGATGGGTAAAAACA 234  
 QY 80 LysThrAspAlaSerValGlyIleSerAsnProSerAsnPheThr---IleProTyrThr 98  
 Db 235 CCATCAGATTCTAGC-----AATACTAATTTCTCAATTTTCTGAAAAAGACTATTTCT 288  
 QY 99 AlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrThrPheAla 117  
 Db 289 TTCAGATATGAACAATCCGTTTGGTTCGTCGAGCAATTTGGTACTCAATG--- 345  
 QY 118 GluGlyThrArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGly 137  
 Db 346 AATGGACCAAGATAGAGTTCGAAAGTATCTCTATGAAACTTTTGTATGTAATAAACCTAG 405  
 QY 138 GlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsn 157  
 Db 406 GGCACACTATAAAACACGACACATGCTACTGTCTTTA-----GATACAGCACCACAA 459  
 QY 158 SerPheThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAsp 177  
 Db 460 AATAGCACTAATGGCGCAGGATTAACATACATCTGTT-----ATCGTAAAAAACGAA 510  
 QY 178 GlyLeuSerIleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnLeu 197  
 Db 511 AATTTACAAATATATCATTAATGTTAAATGCGTGTATGATATCATCTCTGATGGAATA 570  
 QY 198 SerIleSerProTyrIleCysGlyAlaGlyValAspAlaIleGluPheAspVal 217  
 Db 571 CCAGTTTCTCATATGATGTCAGGATTCGACTGATGCTAGTGTCTAGTAAATTAATGCT 630  
 QY 218 LeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsn 237  
 Db 631 ACMAATCTTAATATCTTATCAAGGAAGCTAGGCATTAAGTACTCAATCAATCTGAA 690  
 QY 238 IleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeu 257  
 Db 691 GCTTCTATCTTTATCGGTGGACATTTCCATAGAGTTATAGGTAAATGAATTTAAA----- 744  
 QY 258 AsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSer----- 273  
 Db 745 -----GATATTGCTACCTTTAAAAATATTACTTCAAAAACAGGAATA 786  
 QY 274 -----AlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGluIleGly 288

```

Db 787 TCTAATCTGGCTTGCATCAGCAACTGTGATGTTTGTCACTTGTGGTATAGAAATTGGA 846
Qy 289 AlaArgLeuThrPhe 293
Db 847 GGAAGGTTGTATTT 861

RESULT 26
AAX34745
ID AAX34745 standard; DNA; 843 BP.
AC AAX34745;
XX
DT 05-JUL-1999 (first entry)
XX
DE DNA encoding OMP-1C protein.
XX
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog; ss.
XX
OS Ehrlichia chaffeensis.
XX
FN WO9913720-A1.
XX
PD 25-MAR-1999.
XX
PF 18-SEP-1998; 98WO-US19600.
XX
PR 19-SEP-1997; 97US-0059353.
XX
PA (OHIS ) UNIV OHIO STATE.
XX
PI Ohashi N, Rikihisa Y;
XX
DR WPI; 1999-254290/21.
XX
DR P-PSDB; AAY06945.
XX
XX
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
XX Ehrlichia canis
XX
XX Disclosure; Fig 5A; 55pp; English.
XX
XX The invention provides isolated outer membrane proteins (OMP) from
XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
XX of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
XX in AAY06943-958. The E. canis proteins form part of the P30 family and
XX consist of proteins shown in AAY06959-970. The proteins and genes are
XX used to detect E. chaffeensis in patients and E. canis in dogs.
XX
XX Sequence 843 BP; 262 A; 145 C; 169 G; 267 T; 0 other;

Alignment Scores:
Pred. NO.: 2.16e-39 Length: 843
Score: 450.50 Matches: 105
Percent Similarity: 54.58% Conservative: 56
Best Local Similarity: 35.59% Mismatches: 117
Query Match: 30.11% Indels: 17
DB: 20 Gaps: 7

US-10-062-624-40 (1-293) x AAX34745 (1-843)
Qy 1 MetAsnLysLeuLysPheThrIleLeuValLeuValCysLeuLeuSerLeu 20
Db 1 ATGAACTGCAAA---AAATTTTATACAACTGCAATGGCAATGCTCTTCTTA 57
Qy 21 ProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyThrGlyLeuTy 40
Db 58 CCTGAATATTACTTCTGAACCAAGTACAGATGAGT---GGCAATTTCTAT 114
Qy 41 IleSerGlyGlnTyThrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60
Db 115 ATTAGTGCAGATACATGCAAGTCTCTCATTTTGGAGTTTCTCTGCCAAGAAGAA 174
Qy 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80

```

```

Db 175 AAAAATCCTACTGTGCGCTTGTATGGTTTGAACAAGATTGGAACGGTGTAGTGCCTCA 234
Qy 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyThrAlaVal 100
Db 235 AGTCATGCTGATCGGACTTTTAATACAAAGT-----TATCTTTTAAA 279
Qy 101 PheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyThrPheAlaGluGly 119
Db 280 TACGAAACAAATCCATTTCTAGGTTTTCAGGAGCTATTTGGTTATTCATCAATGGGT---GGT 336
Qy 120 ThrArgValGluIleGluGlySerTyGluCluPheAspValLysAsnProGlyGlyTy 139
Db 337 CCAAGAATAGAGTTTGAAGTGTCTATGAACAATTTGACGTGNAAAAATCAAGGTGGTAA 396
Qy 140 ThrLeuSerAspAlaTyArgTyThrPheAlaLeuAlaArgGluMetLysGlyAsnSerPhe 159
Db 397 TACAAAATGATGCTCAGATACTGTGCTTAGATCGTAAAGCAGCAGCAGCAATATGCC 456
Qy 160 ThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeu 179
Db 457 ACAGCTAGTCAC-----TACGTGCTACTAAAAAATGAAGGACTA 495
Qy 180 SerIleIleSerValIleValAsnValCysTyThrAspPheSerLeuAsnAsnLeuSerIle 199
Db 496 CTTCATATATCATTATGTTGAATGCATGCTATGACGTAGTAGGAAGAAATACCTTTTC 555
Qy 200 SerProTyThrIleCysGlyGlyAlaGlyValaAspAlaIleGluPhePheAspValLeuHis 219
Db 556 TCTCTTACATATGTGCGAGTGTGGTACCGATTAATATCCATGTTTGAAGCTATATAAC 615
Qy 220 IleLysPheAlaTyThrGlnSerLysLeuGlyIleAlaTyThrSerLeuProSerAsnIleSer 239
Db 616 CCTAAATTTCTTATCAAGGAAGCTAGGTTGAGTTACTCTATAAAACCCAGAGCTTCT 675
Qy 240 LeuPheAlaSerLeuTyThrHisLysValMetGlyAsnGlnPheLysAsnLeuAsn--- 258
Db 676 GTCTTTGTGTGGACATTTTTCATAAAGTTGACAGTAAATGAATTCAGGACATTTCTACT 735
Qy 259 ValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeu 278
Db 736 CTTAAAGGTTTGTACACCATCATCTGCAGCTACTCCAGACTTAGCAACAGTAACACTG 795
Qy 279 AsnIleGlyTyThrPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293
Db 796 AGTGTGTGTCACTTTGGAGTAGAAGTTGGAGGAAGATTTAACTTC 840

RESULT 27
AAX01295
ID AAX01295 standard; DNA; 2037 BP.
XX
XX AAX01295;
XX
DT 12-OCT-2000 (first entry)
XX
DE Ehrlichia canis immunoreactive protein genes Eca28SA2 and Eca28SA3.
XX
XX Homologous mature 28-kDa protein gene; Eca28SA2; Eca28SA3; vaccine;
XX p28 gene; polymorphic multiple gene family; immunoprotective antigen;
XX antibacterial; canine ehrlichiosis; canine tropical pancytopenia;
XX tick-borne rickettsial disease; serodiagnosis; immunoreactive; ds.
XX
XX Ehrlichia canis.
XX
XX Key Location/Qualifiers
XX CDS 1..852
XX /tag= a
XX /product= "Eca28SA2 protein (30-kDa)"
XX 853..1194
XX /tag= b
XX /notes= "Intergenic non-coding region NC2"
XX 1195..2037
XX /tag= c

```

1414	-----GGTACAAATCTAACTCTCTCCAGAAAATATATTACACAGTTCCAAAATATTTCG	1467
99	AlaValPheGlnAspAsnSer---ValIlePheAsnGlyThrIleGlyTyrThrPheAla	117
1468	TTTAAATACGAAACCAACCCATCTTGGGTTTGCAGGAGCTATTGGTTTATTCAATGGGT	1527
118	GluGlyThrArgValGluIleGluGlySerTyrGluGluPheAspValIysAsnProGly	137
1528	-----GGCCCAAGAATAGAACTTGAAGTTCTGTACAGACATTGATGTGAAANAATCAGAAC	1584
138	GlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaAargGluMetIysGlyAsn	157
1585	AATAATTATAAGAACGGCGCACACAGATACTGCTGTTTATCTCATCAT-----AGT	1635
158	SerPheThrProIysGluIysValSerAsnSerIlePheHisThrValMetArgAsnAsp	177
1636	TACGCAACCAAGCATGTCCTCCCAAGTAAACAAATTT-----GTTTTCTTTAAAAAATGAA	1689
178	GlyLeuSerIleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeu	1977
1690	GGGTTAAATGCATTATCATTTATGATAAATGCATGCTATGACATAATAATTGAAGGAATG	1749
198	SerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPheAspVal	217
1750	CTTTTTCACCTTATATTGTGCAGGTGTGGTACTGATGTTGTTTCCATGTTTGAAGCT	1809
218	LeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsn	237
1810	ATAATCTCTAAATTTCTTACCAGGAAAACCTAGGATTAGGTTATAGTATATAGTTACAGAA	1869
238	IleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeu	257
1870	GCCTCTGTTTTTATCGTGGACACTTTCACAGAGTCATAGGTAAATGAATTTAGACATC	1929
258	AsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThr	277
1930	CTGCTATG---GTTCTTAGTGGATCAAAATCTTCCAGAAAACCAATTTGCATAGTAAACA	1986
278	LeuAsnIleGlyTyrPheGlyGlyGluIleGlyValaAargLeuThrPhe	293
1987	CTAAATGTGTGCACCTTTCGCATAGAACTTGGAGGAAGATTTTAACTTC	2034
RESULT 28		
AAAD01294		
ID	AAAD01294 standard; DNA; 840 BP.	
XX	AAAD01294;	
XX	12-OCT-2000 (first entry)	
XX	Ehrlichia canis immunoreactive protein Eca28SA3 DNA.	
XX	Homologous 28-kDa protein gene; Eca28SA3; immunoreactive; vaccine;	
KW	p28 gene; polymorphic multiple gene family; immunoprotective antigen;	
KW	antibacterial; canine ehrlichiosis; canine tropical pancytopenia;	
KW	tick-borne rickettsial disease; serodiagnosis; ds.	
XX	Ehrlichia canis.	
OS		
XX	Key	Location/Qualifiers
FH	CDS	1..840
FT		/tag= a
FT		/product= "Eca28SA3 protein (30-kDa)"
FT		/note= "Does not include stop codon"
FT		/partial
FT	sig_peptide	1..69
FT		/tag= b
FT		70..840
FT	mat_peptide	
FT		/tag= c
FT		/product= "Mature Eca28SA3 protein (28-kDa)"
XX		
XX	WO200032745-A2.	
XX		



CC represent the 28-kDa antigen coding sequences and PCR primers of the  
 CC invention.

XX Sequence 840 BP; 282 A; 137 C; 150 G; 271 T; 0 other;

#### Alignment Scores:

Pred. No.: 5,848-39 Length: 840  
 Score: 446.50 Matches: 105  
 Percent Similarity: 55.82% Conservative: 58  
 Best Local Similarity: 35.96% Mismatches: 110  
 Query Match: 29.85% Indels: 19  
 DB: 24 Gaps: 10

US-10-062-624-40 (1-293) x ABK68854 (1-840)

QY 6 LysPheThrIleAlaAsnThrValLeuValCysLeuLeu---SerLeuProAsnIleSer 24  
 DB 10 AAAAAAATTCCTATACCAACATGCAATTAATGCTACTAATGCTCTCAAGCATATCT 69  
 QY 25 SerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyrIleSerGlyGln 44  
 DB 70 TTTTCTGATATACAGACATACACTGGTAGCTTC-----TACATCAGTGGAAAA 123  
 QY 45 TyrLysProSerValSerValPheSerAsnPheSerValLysGluThrAsnValIleThr 64  
 DB 124 TATGTACCAAGTGTTCACATTTGGTGTCTCAGCTAAGAGAAAGAACTCACT 183  
 QY 65 LysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLysThrAspAlaSer 84  
 DB 184 GTTGAGTGTTCCTCCAGAAATATATTCACAGTTCAAAATATTCGTTTAAATACGAA 225  
 QY 85 ValGlyIleSerAsnProSerAsn---PheThrIlePro---TyrThrAlaValPheGln 102  
 DB 226 ATATCTAATCTCTCCAGAAATATATTCACAGTTCAAAATATTCGTTTAAATACGAA 285  
 QY 103 AspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrPheAlaGluGlyThrArg 121  
 DB 286 AACACCCATTCCTAGGTTTCAGGAGCTATGTTATTCATGGGT---GGCCCAAGA 342  
 QY 122 ValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThrLeu 141  
 DB 343 ATAGAACTTGAAGTTCGTACGAGACATTCGATGTGAAATATCAGAACAAATATATAAG 402  
 QY 142 SerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrPro 161  
 DB 403 AACGGCGCACAGATCTGCTTATCTCATCAT------AGTTCAGCAACAAGC 453  
 QY 162 LysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIle 181  
 DB 454 ATGTCTCCGCAAGTAACAATTT-----GTTTCTTAAATAATCAAGGTTAATTGAC 507  
 QY 182 IleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSerPro 201  
 DB 508 TTATCATTTATGATAAATGCATGCTATGATGATGATGATGATGATGATGATGATGATGAT 567  
 QY 202 TyrIleCysGlyGlyValGlyValAspAlaIleGluPheAspValIleHisIleLys 221  
 DB 568 TATATTTGCGAGGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 627  
 QY 222 PheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPhe 241  
 DB 628 ATTTCTTACCAAGGAAACATAGGATAGGATAGGATAGGATAGGATAGGATAGGATAGGAT 687  
 QY 242 AlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGlnHis 261  
 DB 688 ATCGGTGGACACTTTTCACAGAGTCTATAGGTAATGAAATTTAGAGACATCCCTGCTATG 744  
 QY 262 ValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIleGly 281  
 DB 745 GTTCCTAGTGGATCAATCTTCAGAAACCAATTTGCAATAGTACATCAATAGTGTGTGT 804  
 QY 282 TyrPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293

DB 805 CACTTTGGCAGTAACTGGAGGAAGATTTAACTTC 840

#### RESULT 30

AAX34762  
 ID AAX34762 standard; DNA; 843 BP.

XX AC AAX34762;

XX 05-JUL-1999 (first entry)

XX DNA encoding P30-2 protein.

XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 KW detection; dog; ss.

XX OS Ehrlichia canis.

XX PN WO9913720-A1.

XX PD 25-MAR-1999.

XX PF 18-SEP-1998; 98WO-US19600.

XX PR 19-SEP-1997; 97US-0059353.

XX PA (OHIS ) UNIV OHIO STATE.

XX PI Ohashi N, Rikihisa Y;

XX DR WPI; 1999-254290/21.

XX P-PSDB; AAY06962.

XX Novel outer membrane proteins from Ehrlichia chaffeensis and

XX Ehrlichia canis

XX Disclosure; Fig 22A; 55pp; English.

XX The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and  
 CC consist of proteins shown in AAY06959-970. The proteins and genes are  
 CC used to detect E. chaffeensis in patients and E. canis in dogs.

XX SQ Sequence 843 BP; 283 A; 136 C; 150 G; 274 T; 0 other;

#### Alignment Scores:

Pred. No.: 7,538-39 Length: 843  
 Score: 445.50 Matches: 105  
 Percent Similarity: 55.82% Conservative: 58  
 Best Local Similarity: 35.96% Mismatches: 110  
 Query Match: 29.78% Indels: 19  
 DB: 20 Gaps: 10

US-10-062-624-40 (1-293) x AAX34762 (1-843)

QY 6 LysPheThrIleAlaAsnThrValLeuValCysLeuLeu---SerLeuProAsnIleSer 24  
 DB 10 AAAAAAATTCCTATACCAACATGCAATTAATGCTACTAATGCTCTCAAGCATATCT 69  
 QY 25 SerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyrIleSerGlyGln 44  
 DB 70 TTTTCTGATATACAGACATACACTGGTAGCTTC-----TACATCAGTGGAAAA 123  
 QY 45 TyrLysProSerValSerValPheSerAsnPheSerValLysGluThrAsnValIleThr 64  
 DB 124 TATGTACCAAGTGTTCACATTTGGTGTCTCAGCTAAGAGAAAGAACTCACT 183  
 QY 65 LysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLysThrAspAlaSer 84  
 DB 184 GTTGAGTGTTCCTCCAGAAACCAATTTGCAATAGTACATCAATAGTGTGTGT 225  
 QY 85 ValGlyIleSerAsnProSerAsn---PheThrIlePro---TyrThrAlaValPheGln 102

```

Db 226 ATATCTAACTCTCTCCAGAAATATATTCACAGTTCAAAATATTCGTTTAAATACGAA 285
Qy 103 AspAsnSer--ValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArg 121
Db 286 AACACCCATCTTAGGCTTTCAGAGGACTATTCGTTATTCATGGGT--GGCCCAAGA 342
Qy 122 ValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThrLeu 141
Db 343 ATAGAATCTGAAGTTCTCTACGAGACATTCGATGTGAAAAATCAGAACAAATATTAAG 402
Qy 142 SerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrPro 161
Db 403 AACGGCGCACAGATCTGCTTTATCTCATCAT-----AGTTCAGCAACAAC 453
Qy 162 LysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIle 181
Db 454 ATGCTCTCCGCAAGTAACAAATTT-----GTTTCTTAAAAAATGAAGGTTAATTCAC 507
Qy 182 IleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSerPro 201
Db 508 TTATCATTTATGATAAATGCATGCTATGACATAATAATGAAGGAATGCCTTTTTCACCT 567
Qy 202 TyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIleLys 221
Db 568 TATATTTGTGAGGTTGGTACTGATGTTGTTCCATGTTTGAAGCTATAAATCCTAAA 627
Qy 222 PheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPhe 241
Db 628 ATTCTTACCAAGGAATACTAGGATTAGTTATAGTATATAGTTTCAGAGCCTCTGTTTTT 687
Qy 242 AlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGlnHis 261
Db 688 ATCGGTGCACACTTTTCACAGATCATAGGTAATGAATTTAGAGACATCCCTGCTATG-- 744
Qy 262 ValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIleGly 281
Db 745 GTTCTAGTGATCAATATCTTCAGAAACCAATTTGCAATAGTAACACTAAATGTGTGT 804
Qy 282 TyrPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293
Db 805 CACITTTGTTAGAACTTGGAGGAGATTAACTTC 840

```

## RESULT 31

AAAX34761  
ID AAX34761 standard; DNA; 924 BP.

AC AAX34761;

XX 05-JUL-1999 (first entry)

XX DNA encoding p30-1 protein.

DE Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;

KW detection; dog; ss.

XX Ehrlichia canis.

OS WO9913720-A1.

XX 25-MAR-1999.

XX 18-SEP-1998; 98WO-US19600.

XX 19-SEP-1997; 97US-0059353.

XX (OHIS ) UNIV OHIO STATE.

XX Ohashi N, Rikihisa Y;

XX WPI, 1999-254290/21.

DR P-P8DB; AAY06961.

XX

PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
PT Ehrlichia canis

XX Disclosure; Fig 21A; 55pp; English.

CC The invention provides isolated outer membrane proteins (OMP) from  
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
CC in AAY06943-958. The E. canis proteins form part of the P30 family and  
CC consist of proteins shown in AAY06959-970. The proteins and genes are  
CC used to detect E. chaffeensis in patients and E. canis in dogs.

XX Sequence 924 BP; 301 A; 158 C; 160 G; 305 T; 0 other;

## Alignment Scores:

Pred. No.: 1,41e-38 Length: 924  
Score: 443.50 Matches: 106  
Percent Similarity: 54.18% Conservative: 56  
Best Local Similarity: 35.45% Mismatches: 110  
Query Match: 29.65% Indels: 27  
DB: 20 Gaps: 8

US-10-062-624-40 (1-293) x AAX34761 (1-924)

```

Qy 2 AsnAsnLysLysLeuLysPheThrIleAsnThrValLeuValCysLeuLeu---SerLeu 20
Db 85 AATATGAATTCGCAAAAAATTCCTATAACAACATGCATTAATATCATTAATGACTCTATT 144
Qy 21 ProAsnIleSerSerSerLysAlaIleAsnAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40
Db 145 CCAAGCATATCTTTTCTGATCTATATA---CAAGATCGTAACATGGTGGTAACTTCTAT 201
Qy 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60
Db 202 ATTAGTGAAGATGATGACCAAGTCTCCACATTTGCTAGTCTCTCAGCTAAAGAGAA 261
Qy 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
Db 262 AGCAAAATCAACTGTTGGATTTTGGATTAATAAATCATGATGGGATGGA----- 309
Qy 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIlePro---TyrThrAla 99
Db 310 -----AGTCCAAATCTTAAGAATAAACAACGCTGACTTTTACTGTTCACAACTATTCGTT 363
Qy 100 ValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGlu 118
Db 364 AGATACGAGAACATCCATTTCTAGGGTTTCAGAGACTATCGTTACTCAATGGGT--- 420
Qy 119 GlyThrArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGlyGly 138
Db 421 GGCCCAAGAATAGAAATTCGAAATATCTTATGAGCATTCGACGTAAGGCTTAATAATATC 480
Qy 139 TyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGlu-----Met 154
Db 481 AATTATCAAAATGACGCGCAGGCTACTGCGCTCTATCTCATCACATCGGACGCAATG 540
Qy 155 LysGlyAsnSerPheThrProLysGluLysValSerAsnSerIlePheHisThrValMet 174
Db 541 GAAGCTGATAAATTT-----GTCTTCTTA 564
Qy 175 ArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysTyrAspPheSerLeu 194
Db 565 AAAAAACGAAGGGTTAATTGACATATCACTTGAATAAATGCGATGTTATGATATAATAAT 624
Qy 195 AsnAsnLeuSerIleSerProTyrIleCysGlyAlaGlyValAspAlaIleGluPhe 214
Db 625 GACAAAGTACTGTTTCTCTCTATATATGCGCAGGTTATGCTACTGATTTGATTTCTATG 684
Qy 215 PheAspValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeu 234
Db 685 TTGGAAGCTACAAGTCTCTAAATTTCTACAGAGAAACTGGGCATATTGATCTCTATT 744
Qy 235 ProSerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPhe 254

```

```

Db      745 AATCGGAAACCTCTGTTTTCATCGTGGGCAATTTCCACAGGATCATAGTAATGAGTTT 804
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      255 LysAsnLeuAsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAla 274
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      805 AGAGATATTCTGCAATAGTACCTAGTAACCTCACTACATAAGTGGACCAAAATTTGCA 864
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      275 ValAlaThrLeuAsnIleGlyTyrPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      865 ACAGTAACACTAAATGTGTGTCATCTTGGTTTGAAGCTGGGAGGAGATTAACTTC 921
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

## RESULT 32

```

RAD01292
ID   RAD01292 standard; DNA; 1607 BP.
XX
AC   RAD01292;
XX
DT   12-OCT-2000 (first entry)
XX
DE   Ehrlichia canis immunoreactive protein Eca28-1 DNA.
XX
KW   Homologous 28-kDa protein gene; Eca28-1; immunoreactive; vaccine;
KW   p28 gene; polymorphic multiple gene family; immunoprotective antigen;
KW   antibacterial; canine ehrlichiosis; canine tropical pancytopenia;
KW   tick-borne rickettsial disease; serodiagnosis; ds.
XX
OS   Ehrlichia canis.
XX
FH   Key
FT   CDS
FT   /tag= a
FT   /product= "Eca28-1 protein (30-kDa)"
FT   sig_peptide
FT   /tag= b
FT   mat_peptide
FT   /tag= c
FT   /product= "Mature Eca28-1 protein (28-kDa)"
XX
PN   WO200032745-A2.
XX
PD   08-JUN-2000.
XX
PF   24-NOV-1999; 99WO-US28075.
XX
PR   30-NOV-1998; 98US-0201458.
PR   03-MAR-1999; 99US-0261358.
XX
PA   (RERE-) RES DEV FOUND.
XX
PI   Walker DH, Yu X, McBride JW;
XX
DR   WPI; 2000-412298/35.
XX
PT   Ehrlichia canis antigens useful for vaccinating against canine
XX   ehrlichiosis in dogs -
XX
PS   Claim 5; Fig 1; 86pp; English.
XX
CC   The patent relates to homologous 28-kilodalton (kDa) protein genes of
CC   Ehrlichia canis, designated Eca28SA1, Eca28SA2, Eca28SA3, Eca28-1 and
CC   Eca28-2. These genes are members of a polymorphic multiple gene family
CC   and contained in a single locus of 5.92 kb. The 28-kDa proteins are
CC   immunoreactive with anti-E. canis serum hence are important
CC   immunoprotective antigens. The protein is useful for vaccinating
CC   against E. canis infections such as canine ehrlichiosis in dogs.
CC   Canine ehrlichiosis, also known as canine tropical pancytopenia, is a
CC   tick-borne rickettsial disease of dogs. Eca28-1 is conserved amongst
CC   different strains of E. canis and hence useful for serodiagnosis of
CC   canine ehrlichiosis. The present sequence is a DNA encoding E. canis
CC   Eca28-1 30-kDa protein which is post-translationally modified to a
CC   mature 28-kDa protein by cleavage of N-terminal signal sequence.
XX
SQ   Sequence 1607 BP; 543 A; 268 C; 250 G; 540 T; 6 other;

```

RESULT 33



ABK68852  
 ID ABK68852 standard; DNA; 1607 BP.  
 XX AC  
 XX ABK68852;  
 XX DT 02-JUL-2002 (first entry)  
 XX DE DNA encoding Ehrlichia canis p28-7.  
 XX KW Ehrlichia canis infection; vaccine; serodiagnostic; gene; p28; ss;  
 XX KW antibacterial.  
 XX OS Ehrlichia canis.  
 XX PN WO200222782-A2.  
 XX PD 21-MAR-2002.  
 XX PF 12-SEP-2001; 2001WO-US28759.  
 XX PR 12-SEP-2000; 2000US-0660587.  
 XX PA (RERE-) RES DEV FOUND.  
 XX PI Walker DH, Yu X, McBride JW;  
 XX WPI; 2002-351882/38.  
 XX P-PSDB; AAU96100.  
 XX PT New recombinant homologous 28 kilodalton immunodominant protein from  
 XX Ehrlichia canis, useful for treating Ehrlichia canis infections -  
 XX Claim 5; Figure 1; 106pp; English.  
 XX The invention relates to a recombinant homologous 28 kDa immunodominant  
 XX protein, p28 (I) of Ehrlichia canis. (I), a 28-kDa antigen preferably  
 XX dispersed in a pharmaceutically acceptable carrier, is useful for  
 XX inhibiting E. canis infection in a subject. (I) is useful in the  
 XX development of vaccines and serodiagnostics that are particularly  
 XX effective for disease prevention and serodiagnosis. ABK68852-ABK68878  
 XX represent the 28-kDa antigen coding sequences and PCR primers of the  
 XX invention.

US-10-062-624-40 (1-293) x ABK68852 (1-1607)

Alignment Scores:  
 Pred. No.: 3.01e-38 Length: 1607  
 Score: 443.50 Matches: 106  
 Percent Similarity: 54.18% Conservative: 56  
 Best Local Similarity: 35.45% Mismatches: 110  
 Query Match: 29.65% Indels: 27  
 DB: 24 Gaps: 8

US-10-062-624-40 (1-293) x ABK68852 (1-1607)

Qy 2 AsnAsnLysLeuLysPheThrIleAsnThrValLeuValCysLeuLeu---SerLeu 20  
 Db 143 AATATGAATTCGAAAGAAATCTTATAACAATGCTTAAATATCATTAATGCTCTATT 202  
 Qy 21 ProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40  
 Db 203 CCAAGCATATCTTTTCTGATACATATA---CAAGATGGTAAACATGGGTGCTTCTAT 259  
 Qy 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60  
 Db 260 ATTAGTGAAGTAGTACCAAGTGTCTCATTTTGGTAGCTTCTCAGCTAAAGAGAA 319  
 Qy 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80  
 Db 320 AGCAAACTCAACTGTTGGAGTTTGGATTAAACATGATGATGGATGGA----- 367  
 Qy 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIlePro---TyrThrAla 99  
 DB: 99

Db 368 -----AGTCCAATCTTAAGATAAACAACGCTGACTTTTACTGTTCCAAACTATTCGTTTC 421  
 Qy 100 ValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGlu 118  
 Db 422 AGATACGAGAACAAATCCATTTCTAGGGTTTGCAGGAGCTATCGGTACTCAATGGGT--- 478  
 Qy 119 GlyThrArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGlyGly 138  
 Db 479 GGCCCAAGATAGAAATTCGAAATATCTTATGAAGCATTCGACGTAAAGAGCTCTAATATC 538  
 Qy 139 TyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGlu-----Met 154  
 Db 539 AATTATCAAAATGACGCGCACAGTACTGCGCTCTATCTCATCACACATCGGAGCCCATG 598  
 Qy 155 LysGlyAsnSerPheThrProLysGluLysValSerAsnSerIlePheHisThrValMet 174  
 Db 599 GAAGCTGATAAATTT-----GTCTTCTTA 622  
 Qy 175 ArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysTyrAspPheSerLeu 194  
 Db 623 AAAAACGAGGGTTAATTGACATATCACTTGCAATAAATGCATGTTATGATATAATAAAT 682  
 Qy 195 AsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhe 214  
 Db 683 GACAAAGTACTGTTCTCTCTATATATGCGCAGGTATTCGCTACTGATTTGATTCTATG 742  
 Qy 215 PheAspValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeu 234  
 Db 743 TTTGAGCTACAAGTCTCTAAATTTCTTACCAAGGAATACTGGGCATTAGTACTCTATT 802  
 Qy 235 ProSerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPhe 254  
 Db 803 AATCCGGAAACCTCTGTTTTCATCGGTGGGCAATTCACAGGATCATAGGTATGAGTTT 862  
 Qy 255 LysAsnLeuAsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAla 274  
 Db 863 AGAGATATTCTCTGCAATGACTGTAAGTAACTACAATAAGTGGACCAACAATTTGCA 922  
 Qy 275 ValAlaThrLeuAsnIleGlyTyrPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293  
 Db 923 ACAGTAACATTAATGTGTGCTCACTTTGGTTAGACTTGGAGGAGATTTAACTTC 979

RESULT 34  
 AAX34747  
 ID AAX34747 standard; DNA; 837 BP.  
 XX AC AAX34747;  
 XX DT 05-JUL-1999 (first entry)  
 XX DB DNA encoding OMP-1E protein.  
 XX KW Outer membrane protein; OMP; Ehrlichia chafeensis; E. canis; P30;  
 XX KW detection; dog; ss.  
 XX OS Ehrlichia chafeensis.  
 XX PN WO9913720-A1.  
 XX PD 25-MAR-1999.  
 XX PF 18-SEP-1998; 98WO-US19600.  
 XX PR 19-SEP-1997; 97US-0059353.  
 XX PA (OHIS ) UNIV OHIO STATE.  
 XX PI Ohashi N, Rikihisa Y;  
 XX WPI; 1999-254290/21.  
 XX P-PSDB; AAY06947.  
 XX Novel outer membrane proteins from Ehrlichia chafeensis and

```

PT Ehrlichia canis
XX
PS Disclosure; Fig 7A; 55pp; English.
XX
CC The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
SQ Sequence 837 BP; 280 A; 130 C; 159 G; 268 T; 0 other;

Alignment Scores:
Pred. No.: 2.03e-38 Length: 837
Score: 441.50 Matches: 111
Percent Similarity: 56.23% Conservative: 56
Best Local Similarity: 37.37% Mismatches: 107
Query Match: 29.51% Indels: 23
DB: 20 Gaps: 12

US-10-062-624-40 (1-293) x AAX34747 (1-837)
QY 1 MetAsnAsnLeuLysPheThrIleAsnThrValLeuValCysLeuSer--- 19
DB 1 ATGAATTGCAAA---AAATTT---TTTATACAACTGCATTTAGTATCATAATGTCCTTT 54
QY 20 LeuProAsnIleSerSerLysAlaIle---AsnAsnAsnAlaLysLysTyrTyrGly 38
DB 55 CTACCTGGAATATCATTTTCTCATCCAGTGCAAGGTGACAAATATTAGTGGTAATTC--- 111
QY 39 LeuTyrIleSerGlyClnTyrLysProSerValSerValPheSerAsnPheSerVallys 58
DB 112 ---TATGTTAGTGGCAAGTATATGCAAGTGTCTCGCATTTTGGCATGTTTCTGCCAAA 168
QY 59 GluThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGlu 78
DB 169 GAAGAAAAAATCCTACTGTTGCTTATGCTTAAACAGATGGGAAGGATAGC 228
QY 79 ThrLysThrAspAlaSerValGlyLysSerAsnProSerAsnPheThrIleProTyrThr 98
DB 229 TCATCAAGTCACAAATGATAATCATTTCAATAACAAGGT-----TATTCA 273
QY 99 AlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrThrPheAla 117
DB 274 TTTAAATATGAAATAAACCATTTTATAGGTTTGCAGGAGCTATTGGTTATTCAATGGGT 333
QY 118 GluGlyThrArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGly 137
DB 334 ---GGTCCAGAGTAGAGTTTGAAGTGTCTATGAACATTTGACGTTAAANAATCAGGT 390
QY 138 GlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgLysGlyAsn 157
DB 391 AATAACTATAAAAAATGATGCTCACAGATACGTGCTTTAGGTCAACAA-----GACAAC 444
QY 158 SerPheThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAsp 177
DB 445 AGCGGAATACCTAAACTAGTAAA-----TAGCTACTGTATAAAGCGGAA 489
QY 178 GlyLeuSerIleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeu 197
DB 490 GGATGCTTGACATATCATTTATGCTAAATGCTATGCTATGATATAATAACAGAGCATA 549
QY 198 SerIleSerProTyrIleCysGlyAlaGlyValAspAlaIleGluPheAspVal 217
DB 550 CCTTGTCTCTACATATGTCAGGTGTGGTACTGATTTAATATCCATGTTTGAAGCT 609
QY 218 LeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsn 237
DB 610 ACAATCTCTAAATTTCTTACCAAGGAAGTAGGTCTAAGTACTCTATAAACCCAGAA 669
QY 238 IleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeu 257

```

670 GCTTCTGCTATTATTGCTGGACATTTTTCATAGAGTGATAGAAACGAATTTAGGACATT 729  
 258 ---AsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAla 276  
 730 CCTACTCTGAAGCAATTTGTTACGTCATCAGTACTCCAGATCTA-----GCAATAGTA 783  
 277 ThrLeuAsnIleGlyTyrPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293  
 784 ACATAAGTGATGTCATTTTGGTAAGTAAGAACTTGGAGGAAGTTTAACCTTC 834

RESULT 35  
 AAC68704  
 ID AAC68704 standard; DNA; 837 BP.  
 XX  
 AC AAC68704;  
 XX  
 DT 02-MAR-2001 (first entry)  
 XX  
 DE Ehrlichia chaffeensis VSA3 gene partial coding sequence.  
 XX  
 KW Ehrlichia chaffeensis; VSA3; variable surface antigen 3; MAP1;  
 KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;  
 KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;  
 KW 3gdorf3; ds.  
 XX  
 OS Ehrlichia chaffeensis.  
 XX  
 PN WO200065063-A2.  
 XX  
 PD 02-NOV-2000.  
 XX  
 PF 21-APR-2000; 2000WO-US10886.  
 XX  
 PR 22-APR-1999; 99US-0130725.  
 XX  
 PA (UYFL) UNIV FLORIDA.  
 XX  
 PI Barbet AF, Bowie MV, Ganta RR, Burrig MJ, Mahan SM, McGuire TC;  
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;  
 XX  
 DR WPI; 2000-679675/66.  
 DR P-PSDB; AAB36187.  
 XX  
 PT New polynucleotides useful as DNA vaccines for conferring immunity to  
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,  
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens  
 PT  
 Claim 4; Page 40; 63pp; English.  
 CC The present sequence shows a high degree of similarity to the major  
 CC antigenic protein 1 (MAP1) gene of Ehrlichia sp. The MAP1 gene may be  
 CC used in a vaccine to protect animals or humans against rickettsial  
 CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,  
 CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response  
 CC protective against the rickettsial pathogen. The nucleic acid vaccines  
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.  
 CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1  
 CC and 3gdorf3 may be used in therapeutic and diagnostic applications. The  
 CC polypeptides are useful for detecting antibodies associated with  
 CC infection by a rickettsial pathogen whilst the polynucleotides may be  
 CC used to detect the presence of rickettsial nucleic acids.  
 XX  
 SQ Sequence 837 BP; 279 A; 130 C; 159 G; 269 T; 0 other;

Alignment Scores:  
 Pred. No.: 2.03e-38 Length: 837  
 Score: 441.50 Matches: 111  
 Percent Similarity: 56.23% Conservative: 56  
 Best Local Similarity: 37.37% Mismatches: 107  
 Query Match: 29.51% Indels: 23  
 DB: 21 Gaps: 12

US-10-062-624-40 (1-293) x AAC68704 (1-837)

Qy 1 MetAsnLysLeuLysPheThrIleAlaSerValLeuValCysLeuLeuSer--- 19  
Db 3 ATGAATTGCAAA---AAATTT---TTTATAACAACTGCATTAGTATCATCTTAATTCCTTT 56  
Qy 20 LeuProAsnIleSerSerLysAlaIle---AsnAsnAlaLysLysTyTyGly 38  
Db 57 CTACCTGGATATCATTTTCTGATCCAGTGCAAGGTGACAATATTAGTGGTAATTC--- 113  
Qy 39 LeuTyRileSerGlyGlnTyRlysProSerValSerValPheSerAsnPheSerValLys 58  
Db 114 ---TATGTTAGTGGCAAGTATATGCCAAGTGCTTCGCATTTTGGCATGTTTTCTGCGCAA 170  
Qy 59 GluThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSeriLeGlu 78  
Db 171 GAGAIAAAAAATCCTACTGTGTCATTTGATGGCTTTAAAACAAGATCGGAAAGGATAGC 230  
Qy 79 ThrLysThrAspAlaSerValGlyLysSerAsnProSerAsnPheThrIleProTyThr 98  
Db 231 TCATCAAGTCACAATGATAATCATTTCAATAACAAGGT-----TATTCA 275  
Qy 99 AlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyThrPheAla 117  
Db 276 TTTAAATATGAAAAATAACCCATTTTATGGGTTTGCAGGAGCTATTTGGTTATTCAATGGGT 335  
Qy 118 GluGlyThrArgValGluIleGluGlySerTyRgluGluPheAspValLysAsnPcOgly 137  
Db 336 ---GGTCCAAGAGTAGAGTTTGAAGTGCTCTATGAAACATTTGACGTTAAANAATCAGGT 392  
Qy 138 GlyTyThrLeuSerAspAlaTyRArgTyRPhelalaLeualaArgGluMetLysGlyAsn 157  
Db 393 AATAACTATAAAAATGATGCTCACAGATCTGTGCTTTAGGTCAACAA-----GACAA 446  
Qy 158 SerPheThrProLysGluLysValSerAsnSeriLePheHisThrValMetArgAsnAsp 177  
Db 447 AGCGGNATACCTAAACTAGTAA-----TACGTACTGTTTAAAGCGAA 491  
Qy 178 GlyLeuSeriLeSerValIleValAsnValCystyrAspPheSerLeuAsnLeu 197  
Db 492 GGATGCTTGACATATCATTTATGCTAAATGCATGCTATGATATAATAAACGAGAGCAT 551  
Qy 198 SerylSerProTyRileCysGlyAlaGlyValaspallieGluPhePheAspVal 217  
Db 552 CCTTTGCTCTTACATATGCGAGGTGTTGGTACTGATTTAATPACCATTGTTGAAGCT 611  
Qy 218 LeuHisIleLysPheAlaTyRGlnSerLysLeuGlyIleAlaTyRserLeuProSerAsn 237  
Db 612 ACAATCCTAAAATTTCTTACCAGGGAAGTTAGTCTAGTTACTCTATAAACCCAGAA 671  
Qy 238 IleSerLeuPheAlaSerLeuTyRHisLysValMetGlyAsnGlnPheLysAsnLeu 257  
Db 672 GCTTCTGTTATTTATTTGGTGGACATTTTCATTAAGGTGATAGGAACGAATTTAGGACATT 731  
Qy 258 ---AsnValGlnHisValalagluLeualaserileProLysIleThrSerAlaValAla 276  
Db 732 CCTACTCTGAAGACATTTGTTAGCTCATCAGTACTCAGATCTA-----GCATAGTA 785  
Qy 277 ThrLeuAsnIleGlyTyRPhedgylGluIleGlyAlaArgLeuThrPhe 293  
Db 786 ACACCTAAGTGTATGTCATTTTGGATAGACTCTGGAGAGGTTTAACTTC 836

RESIT.T 36

RESOLI 30  
AAN01293

AAD01293  
ID AAD01293 standard: DNA: 849 bp.

AC AAD01293:

AC  
XX  
DT

DE Ehrlichia canis immunoreactive protein ECa28SA2 DNA.

Homologous 28-kDa protein gene; ECa28SA2; immunoreactive; vaccine; KW

KW	KW	KW	XX	OS	XX	FH	FT	FT	FT	FT	XX	PX	PN	XX	PD	XX	PF	XX	PR	PR	XX	PA	XX	PI	XX	DR	DR	XX	PT	PS	XX	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	SQ		Alii	Pre	Sec	Pen	Bes	Que	DB		US -	QY	Dd	QY	Dd	QY	Dd	QY	
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	--	------	-----	-----	-----	-----	-----	----	--	------	----	----	----	----	----	----	----	--

p28 gene; polymorphic multiple gene family; immunoprotective antigen; antibacterial; canine ehrlichiosis; canine tropical pancytopenia; tick-borne rickettsial disease; serodiagnosis; ds.

Key	Location/Qualifiers
CDS	1..849
	/*tag= a
	/product= "Eca28SA2 protein"
	/note= "Does not include sto
	/partial

WO200032745-A2.

08-JUN-2000.

24-NOV-1999: 99WO-US28075.

30-NOV-1998: 98US-0201458

03-MAR-1999; 9905-0261358.

(KERE-) RES DEV FOUND.

Walker DH, Yu X, McBride JW;

WPI; 2000-412298/  
B-PCPB: AY71478

14-00000

ehrlichiosis in d

Claim 5; Page 65-

The patent relates

Eca28-2. These ge

immunoreactive wi

against *E. canis*

tick-borne ricket

canine ehrlichiosis

mature 28-kDa pro

Sequence 849 BP;

**Comment Scores:**

..

Local Similarity:

100

0-062-624-40 (1-29

# 1 MetAsnAs

1 ATGAATTG

21 ProAsn11

58 CCTAATGT

41 Tlesergl

115 ATATCAGG

61 AsnValT11

```

Db      175 AAAAAAAAAAGCACTGTAGTATATGCGTTAAAGAAAC-----TGGGCA 219
Qy      81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIle---ProTyrThrAla 99
Db      220 GGAGATGCAATATCTAGTCAAGTCCAGATGATAATTTTACCATTCGAAATTTACTCATTC 279
Qy      100 ValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGlu 118
Db      280 AAGTATGCAAGCAACAAGTTTGTAGGTTTGCAGTAGCTATTGGTTACTCGATAGGC--- 336
Qy      119 GlyThrArgValGluLeuGlySerTyrGluGluPheAspValIleAsnProGlyGly 138
Db      337 AGTCCAAGAAATAGAGTTGAGATGCTTATGAAGCATTTTGTGAAATATCCAGGTGAT 396
Qy      139 TyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetIleGlyAsnSer 158
Db      397 AATTACAAAACGGTCTTACAGGTATTGCTTTATCTCATCAAGATGATCGGATGAT 456
Qy      159 PheThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGly 178
Db      457 -----GACATGACTAGTGCAACTGACAAATTTGTATATTATTAATGAAGCA 504
Qy      179 LeuSerIleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSer 198
Db      505 TTACTTAACATATCATTTTATGACAAACATATGTTATGAACAGCAGCAAAATATATCCT 564
Qy      199 IleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeu 218
Db      565 CTCTCTCTTACATATGTCAGGTATTGCTAGTATTATTAATTCACATGTTTGAACACTACA 624
Qy      219 HisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIle 238
Db      625 CATCTTAAATTTCTTATCAAGGAAGCTGAGGTTGGCTTACTTCTGTAAGTCAGAGTCT 684
Qy      239 SerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsn 258
Db      685 TCGGTTCTTTGGTATATATTTTCAAAAAATTAATAATTAATAAGTTTAAATAATGTCCA 744
Qy      259 ValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeu 278
Db      745 GCCATGGTACCTATTAACTCAGACGAGATAGTAGGACCACAGTGTGCAACAGTAACATTA 804
Qy      279 AsnIleGlyTyrPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293
Db      805 AATGATGTACTTTGGATTAGAACTTGGATGTAGGTTCAACTTC 849

RESULT 37
ID      ABK68853 standard; DNA; 849 BP.
AC      ABK68853;
XX      02-JUL-2002 (first entry)
DT      DNA encoding Ehrlichia canis p28-5.
DE      Ehrlichia canis infection; vaccine; serodiagnostic; gene; p28; ss;
KW      antibacterial.
XX      Ehrlichia canis.
XX      WO200222782-A2.
XX      21-MAR-2002.
XX      12-SEP-2001; 2001WO-US28759.
XX      12-SEP-2000; 2000US-0660587.
XX      (RERE-) RES DEV FOUND.
XX      Walker DH, Yu X, McBride JW;

```

```

XX      WPI; 2002-351882/38.
DR      P-PSDB; AAU96101.
XX      New recombinant homologous 28 kilodalton immunodominant protein from
PT      Ehrlichia canis, useful for treating Ehrlichia canis infections -
XX      Claim 5; Figure 7; 106pp; English.
PS      The invention relates to a recombinant homologous 28 kDa immunodominant
CC      protein, p28 (I) of Ehrlichia canis. (I), a 28-kDa antigen preferably
CC      dispersed in a pharmaceutically acceptable carrier, is useful for
CC      inhibiting E. canis infection in a subject. (I) is useful in the
CC      development of vaccines and serodiagnostics that are particularly
CC      effective for disease prevention and serodiagnosis. ABK68852-ABK68878
CC      represent the 28-kDa antigen coding sequences and PCR primers of the
CC      invention.
XX      SQ      Sequence 849 BP; 293 A; 131 C; 152 G; 273 T; 0 other;
Alignment Scores:
Pred. No.: 2,34e-38 Length: 849
Score: 441.00 Matches: 109
Percent Similarity: 52.54% Conservative: 46
Best Local Similarity: 36.95% Mismatches: 126
Query Match: 29.48% Indels: 14
Db: 24 Gaps: 7
US-10-062-624-40 (1-293) x ABK68853 (1-849)
Qy      1 MetAsnLysLeuLysPheThrIleAsnThrValLeuValCysLeuSerLeu 20
Db      1 ATGAATTTGTAATAAAGTTTTCACA---ATAAGTCATGATGATCATCTCATTACTCTCTA 57
Qy      21 ProAsnIleSerSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40
Db      58 CCTAATGCTCATCTCTAACCAGTATATGGTAAACAGT---ATGATGGTAATTTTAC 114
Qy      41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60
Db      115 ATATCAGAAAGTACATGCCAAGTGTCTCTCATTTTGGAAATTTTTCAGCTGAAGAAGAG 174
Qy      61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
Db      175 AAAAAAAGACACTGTAGTATATGCTTAAAGAAAC-----TGGGCA 219
Qy      81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIle---ProTyrThrAla 99
Db      220 GGAGATGCAATATCTAGTCAAGTCCAGATGATAATTTTACCATTCGAAATTTACTCATTC 279
Qy      100 ValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGlu 118
Db      280 AAGTATGCAAGCAACAAGTTTGTAGGTTTGCAGTAGCTATTGGTTACTCGATAGGC--- 336
Qy      119 GlyThrArgValGluLeuGlySerTyrGluGluPheAspValIleAsnProGlyGly 138
Db      337 AGTCCAAGAAATAGAGTTGAGATGCTTATGAAGCATTTTGTGAAATATCCAGGTGAT 396
Qy      139 TyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetIleGlyAsnSer 158
Db      397 AATTACAAAACGGTCTTACAGGTATTGCTTTATCTCATCAAGATGATCGGATGAT 456
Qy      159 PheThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGly 178
Db      457 -----GACATGACTAGTGCAACTGACAAATTTGTATATTATTAATGAAGCA 504
Qy      179 LeuSerIleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSer 198
Db      505 TTACTTAACATATCATTTTATGACAAACATATGTTATGAACAGCAGCAAAATATATCCT 564
Qy      199 IleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeu 218
Db      565 CTCTCTCTTACATATGTCAGGTATTGCTAGTATTATTAATTCACATGTTTGAACACTACA 624

```



PN WO200222782-A2.  
 XX  
 PD 21-MAR-2002.  
 PF 12-SEP-2001; 2001WO-US28759.  
 XX  
 PR 12-SEP-2000; 2000US-0660587.  
 XX  
 PA (RERE-) RES DEV FOUND.  
 XX  
 PI Walker DH, Yu X, McBride JW;  
 XX  
 DR WPI: 2002-351882/38.  
 DR P-PSDB; AAU96118.  
 XX  
 PT New recombinant homologous 28 kilodalton immunodominant protein from  
 PT Ehrlichia canis, useful for treating Ehrlichia canis infections -  
 XX  
 PS Claim 5; Figure 16; 106pp; English.  
 XX  
 CC The invention relates to a recombinant homologous 28 kDa immunodominant  
 CC protein, p28 (I) of Ehrlichia canis. (I), a 28-kDa antigen preferably  
 CC dispersed in a pharmaceutically acceptable carrier, is useful for  
 CC inhibiting E. canis infection in a subject. (I) is useful in the  
 CC development of vaccines and serodiagnostics that are particularly  
 CC effective for disease prevention and serodiagnosis. ABK68852-ABK68878  
 CC represent the 28-kDa antigen coding sequences and PCR primers of the  
 CC invention.  
 XX  
 SQ Sequence 813 BP; 232 A; 106 C; 171 G; 304 T; 0 other;

## Alignment Scores:

Pred. No.: 66-38 Length: 813  
 Score: 437.00 Matches: 110  
 Percent Similarity: 52.38% Conservative: 44  
 Best Local Similarity: 37.41% Mismatches: 110  
 Query Match: 29.21% Indels: 30  
 DB: 24 Gaps: 8

US-10-062-624-40 (1-293) x ABK68878 (1-813)

QY 1 MetAsnAsnLysLeuLysPheThrIleAsnThrValLeuValCysLeuLeuSerLeu 20  
 Db 1 ATGAATTACAAA---AGATTGGTGTAGGTGTACGCTGAGTACATTTGTTTCTTCA 57  
 QY 21 ProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40  
 Db 58 TCTGATGGTGTCTTCTCATGCA-----AATTTTCTGAAGGGAGGAGGACTTTAT 111  
 QY 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60  
 Db 112 ATAGGTAGTCAGTATAAAGTTGGTATTCCTCAATTTTAGTAATTTTCAGCTGAAGAACA 171  
 QY 61 ---AsnValIleThrLysAsnLeuLeuAlaLeuLysLysAspValAspSerIleGluThr 79  
 Db 172 ATTCCTGGTATTACAAAAGATTTTGGTGTAGGTCTTGATAAGTCTGAGATAATACT 231  
 QY 80 LysThrAspAlaSerValGlyLysSerAsnProSerAsnPheThrIleProTyrThrAla 99  
 Db 232 CAC-----AGCAATTTTACACGATCATATGACCCT 261  
 QY 100 ValPheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGly 119  
 Db 262 ACTTATGCAAGCAGTTTTCGAGGTTTATGGGTATCATTTGGATAT---TATGTTAATGAC 318  
 QY 120 ThrArgValGluLeuGluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyr 139  
 Db 319 TTTAGGTTAGATTGAAAGGTTCTTATGAGATTTTGAACCTGAAGACAAATGGTACCCT 378  
 QY 140 ThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMet-----LysGlyAsn 157  
 Db 379 GAGATAGCCAAAGTACAAATTTTGTCTCGAAATGCTCAAAATAGTGATAT 438

QY 158 SerPheThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAsp 177  
 Db 439 AAGTTT-----ATAGTACTAGAGATAAC 462  
 QY 178 GlyLeuSerIleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeu 197  
 Db 463 GCGGTGGTGCACAGTCTCTTAATGTAATGTTTGTATGATATTGCTAGTGGTAGTATT 522  
 QY 198 SerIleSerProTyrIleCysGlyAlaGlyValAspAlaIleGluPhePheAspVal 217  
 Db 523 CCTTTAGCACCTTATATGTGTGCTGGTGGTGCAGATTATATAAAGTTTTTAGGTATA 582  
 QY 218 LeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsn 237  
 Db 583 TCATTGCTAAGTTTCTTATCAAGTTTGGTGTCAACTACCTCTCTAAATGTTAAT 642  
 QY 238 IleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeu 257  
 Db 643 ACTATGTTGTTTGGTGGGGTTATTACCATAAGGTTGTAGGTGATGAGCATGAGAGATA 702  
 QY 258 AsnVal---GlnHisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAla 276  
 Db 703 GAAATAGTCTTACCCTTACTGTCATTATCTGACGTTCTAGAACTACTTCAGCTTCTGCT 762  
 QY 277 ThrLeuAsnIleGlyTyrPheGlyGlyGluIleGlyAlaAig 290  
 Db 763 ACTTTAAATACTGATTATTGTTGGTGGAGATTGGAATTAGA 804

## RESULT 40

AAK34743  
 ID AAK34743 standard; DNA; 846 BP.

XX AAK34743;

DT 05-JUL-1999 (first entry)

DE DNA encoding OMP-1 protein.

XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;

KW detection; dog; ss.

XX Ehrlichia chaffeensis.

XX WO9913720-A1.

XX 25-MAR-1999.

XX 18-SEP-1998; 98WO-US19600.

XX 19-SEP-1997; 97US-0059353.

XX (OHIS ) UNIV OHIO STATE.

XX Ohashi N, Rikihisa Y;

XX WPI: 1999-254290/21.

DR P-PSDB; AAY06943.

XX Novel outer membrane proteins from Ehrlichia chaffeensis and  
 PT Ehrlichia canis

XX Disclosure; Fig 3A; 55pp; English.

XX The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 CC in AAY06943-958. The E. canis proteins form part of the p30 family and  
 CC consist of proteins shown in AAY06959-970. The proteins and genes are  
 CC used to detect E. chaffeensis in patients and E. canis in dogs.

XX Sequence 846 BP; 281 A; 143 C; 164 G; 258 T; 0 other;

Alignment Scores:

Pred. No.: 1.34e-37 Length: 846  
Score: 434.00 Matches: 109  
Percent Similarity: 54.25% Conservative: 57  
Best Local Similarity: 35.62% Mismatches: 102  
Query Match: 29.01% Indels: 38  
DB: 20 Gaps: 13

Search completed: July 3, 2003, 22:38:26  
Job time : 216 secs

US-10-062-624-40 (1-293) x AAX34743 (1-846)

```
Qy 1 MetAsnLeuLysLeuLysPheThrIleAsnThrValLeuValCysLeuLeu---Ser 19
Db 1 ATGAATTACAAAAGTTTC-----ATAACAGTGCATTGATATCATTAATCTCTCT 54

Qy 20 LeuProAsnIleSerSerSer-----LysAlaIleAsnAsnAlaLysLys 35
Db 55 CTACCTGGAGTATCATTTTCGACCCAGCAGGTAGTGTATTAAACGGTAAT----- 105

Qy 36 TyrTyrGlyLeuTyrIleSerGlyGlnTyrIlypSerValSerValPheSerAsnPhe 55
Db 106 -----TTCTACATCAGTGGAAATACATGCCAAGTCTTCGCATTTTGGAGTATTC 156

Qy 56 SerValLysGluThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValasp 75
Db 157 TCTGCTAAGGAAGAAAGAAATACACAGTTCGGAGTGTTCGACTGGACCAAAATGGGAC 216

Qy 76 SerIleGluThrLysThrAspAlaSerValGlyIleSerAsnProSerAsn---PheThr 94
Db 217 GGA-----AGCGCAATATCCAACTCTCCCCCAACGATGTATTCACT 258

Qy 95 IlePro---TyrThrAlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIle 112
Db 259 GTCTCAAAATTATTCATTTAAATATGAAACCAACCCGTTTTTATGGTTTTGCAGGAGCTATT 318

Qy 113 GlyTyrThrPheAlaGluGlyThrArgValGluIleGluGlySerTyrGluGluPheAsp 132
Db 319 GGTACTCAATG---GATGGTCCAGATAGAGCTTGAAGTATCTTATGAACATTTCAT 375

Qy 133 ValLysAsnProGlyGlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArg 152
Db 376 GTAAAAAATCAAGGTAAATATAAGAAATGAAGCACATAGATATTGGTCTCTATCCCAT 435

Qy 153 GluMetLysGlyAsnSerPheThrProLysGluLysValSerAsnSerIlePheHisThr 172
Db 436 -----AACTCAGCAGCAGACATGATAGTGCAGTAATAATTTT-----GTC 477

Qy 173 ValMetArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysTyrAspPhe 192
Db 478 TTCTAAAAAATGAAGGATTACTTGACATATCATTTATGCTGAACGATGCTATGACGTA 537

Qy 193 SerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValaspAlaIle 212
Db 538 GTAGCGGAAGGCATACCTTTTCTCTTATATATGCGCAGGTATCGGTACTGATTAGTA 597

Qy 213 GluPhePheAspValLeuHisIleIlypPheAlaTyrGlnSerLysLeuGlyIleAlaTyr 232
Db 598 TCCATGTTTGAAGCTACAAATCCCTAAAAATTTCTTACCAAGGAAAGTTAGGTTTAAGCTAC 657

Qy 233 SerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsn 252
Db 658 TCTATAGCCAGAGCTTCTGTGTTTATGCTGGGCACCTTCATAGGTAATAGGGAAAC 717

Qy 253 GlnPheLysAsnLeuAsnVal-----GlnHisValAlaGluLeuAlaSer 267
Db 718 GAATTTAGAGATATTCCTACTATAATACCTACTGATCAACACTTGCAGGAAAGGAAAC 777

Qy 268 IleProLysIleThrSerAlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyLeu 287
Db 778 TACCCT-----GCAATAGTAATACTGGATGTATGCCACTTTGGAATAGAACTT 825

Qy 288 GlyAlaArgLeuThrPhe 293
Db 826 GGAGGAAGGTTTGTATTC 843
```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 10:21:10 ; Search time 23 Seconds  
(without alignments)  
528.372 Million cell updates/sec

Title: US-10-062-624-40

Perfect score: 293

Sequence: 1 MNKKLKTIIITVLVCLLSL.....AVATLNIGYGGIGARLTF 293

Scoring table:

OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	2.7	166	1 RL10_NEIMA	Q91qp7 neisseria m
2	8	2.7	1324	1 MSH6_ARATH	O04716 arabidopsis
3	7	2.4	109	1 K2M1_SHEEP	P02539 ovis aries
4	7	2.4	196	1 VATL_DICDI	P54642 dictyosteli
5	7	2.4	292	1 GLXR_ECOLI	P77161 escherichia
6	7	2.4	293	1 FPG_BRUME	Q8yed2 brucella me
7	7	2.4	303	1 HEMZ_CAWJE	Q9pi08 campylobact
8	7	2.4	368	1 IDH2_KLJLA	O84230 kluyveromyc
9	7	2.4	491	1 K2M2_SHEEP	P15241 ovis aries
10	7	2.4	493	1 SECY_ARCFU	O28377 archaeoglob
11	7	2.4	604	1 BAP3_YEAST	P41815 saccharomyc
12	7	2.4	628	1 YKP5_YEAST	P36056 saccharomyc
13	7	2.4	629	1 K2C3_HUMAN	P12035 homo sapien
14	7	2.4	863	1 GLND_HAEIN	P43919 haemophilus
15	7	2.4	900	1 FOX2_YEAST	Q02207 saccharomyc
16	7	2.4	1025	1 YEG0_ECOLI	P76399 escherichia
17	7	2.4	1046	1 SRCC_LACLA	Q8cfz0 lactococcus
18	7	2.4	3587	1 SRP2_BACSU	Q04747 bacillus su
19	6	2.0	55	1 A70A_DROMA	O18666 drosophila
20	6	2.0	55	1 A70A_DROSE	O18417 drosophila
21	6	2.0	80	1 PGTI_SALTY	P41784 salmonella
22	6	2.0	88	1 RL15_PINTH	P41648 pinus thunb
23	6	2.0	96	1 YL34_ARCFU	O28146 archaeoglob
24	6	2.0	107	1 QACG_STASP	O87866 staphylococ
25	6	2.0	107	1 Y098_ARCFU	O30138 archaeoglob
26	6	2.0	108	1 KVIH_HUMAN	P01600 homo sapien
27	6	2.0	109	1 KV3F_HUMAN	P01624 homo sapien
28	6	2.0	111	1 V441_BPMD2	O64236 mycobacteri
29	6	2.0	113	1 DAD1_XENLA	P46967 xenopus lae
30	6	2.0	117	1 NU3M_PARLI	P12774 paracentrot
31	6	2.0	118	1 RL20_ANASP	Q8yrl8 anabaena sp
32	6	2.0	119	1 YN8E_YEAST	P53726 saccharomyc
33	6	2.0	128	1 VG39_BPMD2	O64230 mycobacteri

34	6	2.0	129	1 KV3H_HUMAN	P04207 homo sapien
35	6	2.0	130	1 Y142_LEUMC	P97117 leuconostoc
36	6	2.0	133	1 HSLR_ECOLI	P45802 escherichia
37	6	2.0	135	1 IL4_BOVIN	P30367 bos taurus
38	6	2.0	135	1 IL4_CEREL	P51744 cervus elap
39	6	2.0	144	1 NUSE_HAEIN	P45150 haemophilus
40	6	2.0	145	1 Y127_MYCGE	P47373 mycoplasma
41	6	2.0	146	1 HBB_LATCH	P23741 latimeria c
42	6	2.0	156	1 VGLL_PRVIF	P52511 pseudorabie
43	6	2.0	156	1 VGLL_PRVKA	P52512 pseudorabie
44	6	2.0	160	1 DYR_MYCPN	P78028 mycoplasma
45	6	2.0	162	1 RL10_BORBU	O51352 borrelia bu
46	6	2.0	163	1 COAD_BORBU	O51645 borrelia bu
47	6	2.0	163	1 GLBZ_CHITH	P23761 chironomus
48	6	2.0	163	1 GLBZ_CHITH	P29245 chironomus
49	6	2.0	164	1 PTRB_KLEPN	P37081 klebsiella
50	6	2.0	181	1 CHMU_ERWHE	P42517 erwinia her
51	6	2.0	183	1 AR21_CAEEL	O9xwv3 caenorhabdi
52	6	2.0	187	1 Y503_TREPA	O83516 treponema p
53	6	2.0	189	1 CTR2_YEAST	P38865 saccharomyc
54	6	2.0	189	1 DYR_CAEEL	Q93341 caenorhabdi
55	6	2.0	190	1 C24A_BISBI	Q95173 b cytochrom
56	6	2.0	190	1 C24A_BOVIN	O46521 bos taurus
57	6	2.0	191	1 C24A_PIG	P26550 s cytochrom
58	6	2.0	193	1 C24A_RABIT	Q95m44 o cytochrom
59	6	2.0	196	1 YCH6_YEAST	P25602 saccharomyc
60	6	2.0	199	1 WRBA_ZYMMO	Q9xbr5 zymonomas m
61	6	2.0	200	1 NUPL_XENLA	P05221 xenopus lae
62	6	2.0	200	1 PCXA_PSEPU	P00436 pseudomonas
63	6	2.0	201	1 OSMY_ECOLI	P27291 escherichia
64	6	2.0	202	1 NP3_RHOPR	Q94733 rhodnius pr
65	6	2.0	209	1 RL4_BORBU	P94268 borrelia bu
66	6	2.0	211	1 YFDX_ECO57	P58403 escherichia
67	6	2.0	211	1 YFDX_ECOLI	P76520 escherichia
68	6	2.0	213	1 HS49_YEAST	Q99181 saccharomyc
69	6	2.0	213	1 PIS_EAT	P70500 rattus norv
70	6	2.0	215	1 RNS6_NICAL	Q40379 nicotiana a
71	6	2.0	222	1 PIMT_CAUCR	Q9a6t6 caulobacter
72	6	2.0	224	1 YHCA_ECOLI	P28722 escherichia
73	6	2.0	229	1 Y268_MYCPN	P75396 mycoplasma
74	6	2.0	230	1 ATP6_ASTPE	Q33823 asterina pe
75	6	2.0	230	1 TOLQ_ECOLI	P05828 escherichia
76	6	2.0	233	1 RNS4_ANTI	Q38717 antirrhinum
77	6	2.0	236	1 LIVF_ARCFU	O28882 archaeoglob
78	6	2.0	238	1 YM34_SULSO	Q97w42 sulfoblob
79	6	2.0	240	1 YFIC_HAEIN	P44702 haemophilus
80	6	2.0	241	1 PSMA_SULSO	Q9uxc6 sulfoblob
81	6	2.0	241	1 YCA2_PLAPA	P25408 plasmodium
82	6	2.0	242	1 PSMA_SULTO	Q97595 sulfoblob
83	6	2.0	247	1 YCP4_YEAST	P25349 saccharomyc
84	6	2.0	249	1 RU2A_ARATH	P43333 arabidopsis
85	6	2.0	250	1 RIPB_LUPCY	P22851 luffa cylin
86	6	2.0	254	1 HISA_AZOB	P26720 azospirillu
87	6	2.0	254	1 YC8A_MERJA	P81318 methanococc
88	6	2.0	255	1 TAM_BRUME	Q8y191 brucella me
89	6	2.0	260	1 1433_LYCES	P93209 lycopersico
90	6	2.0	260	1 143C_TOBAC	P93343 nicotiana t
91	6	2.0	261	1 OMPY_CHLTR	O84376 chlamydia t
92	6	2.0	265	1 FABI_PSEAE	Q9zfe4 pseudomonas
93	6	2.0	266	1 KSGA_MYCCA	P43038 m dimethyla
94	6	2.0	266	1 THID_ECOLI	P76422 escherichia
95	6	2.0	267	1 DKGB_ECOLI	Q8x7z7 escherichia
96	6	2.0	267	1 DKGB_ECOLI	P30863 escherichia
97	6	2.0	267	1 DKGB_SALTY	Q8z988 salmonella
98	6	2.0	267	1 DKGB_SALTY	Q8z987 salmonella
99	6	2.0	268	1 PLSC_MYCGE	Q94002 mycoplasma
100	6	2.0	269	1 TRPA_BACST	P19867 bacillus st
101	6	2.0	272	1 ALFI_AERPE	Q9y990 aeropyrum p
102	6	2.0	273	1 TYRO_CANFA	P54834 canis fami
103	6	2.0	273	1 TYRO_COTJA	Q08410 coturnix co
104	6	2.0	273	1 TYRO_TRISI	P55026 trionyx ein
105	6	2.0	274	1 SACT_YEAST	P17121 saccharomyc
106	6	2.0	281	1 YWFL_BACSU	P39648 bacillus su

107	6	2.0	283	1	YOP2_CAEEL	Q09303 caenorhabdi	180	1	CYB_SPHTV	P34875 sphyrna tib
108	6	2.0	291	1	MSAB_TREPA	O83641 t peptide m	181	1	GRPR_HUMAN	P30550 homo sapien
109	6	2.0	291	1	YA38_PYRHO	O58759 pyrococcus	182	1	GRPR_RAT	P32500 rattus norv
110	6	2.0	292	1	EFTS_BACHD	O9ka64 bacillus ha	183	1	K2CD_HUMAN	P48667 homo sapien
111	6	2.0	292	1	EFTS_BACSU	P80700 bacillus su	184	1	TR15_MYRRO	O13489 myrothecium
112	6	2.0	292	1	YK4_YEAST	Q02202 saccharomyc	185	1	METL_ARATH	P17562 arabidopsis
113	6	2.0	293	1	TVRC_ZYMMO	O04983 zymomonas m	186	1	SETB_ECOLI	P33026 escherichia
114	6	2.0	293	1	Y698_CHLMU	O9fjx8 chlamydia m	187	1	SETB_SALTY	P33027 salmonella
115	6	2.0	299	1	YKXC_CYAPA	P31606 cyanophora	188	1	SOTA_ERWCH	Q963k0 erwinia chr
116	6	2.0	303	1	ARAL_STRAT	Q03320 streptomyce	189	1	SETC_ECOLI	P31436 escherichia
117	6	2.0	307	1	YER3_YEAST	P40056 saccharomyc	190	1	KIME_HUMAN	Q03426 homo sapien
118	6	2.0	309	1	HPRK_MYCPU	Q98p11 mycoplasma	191	1	BR53_CAVPO	P35371 cavia porce
119	6	2.0	309	1	MOCB_RHIME	P49308 rhizobium m	192	1	BR53_HUMAN	P32247 homo sapien
120	6	2.0	310	1	DHB3_HUMAN	P37058 homo sapien	193	1	BR53_MOUSE	O54798 mus musculu
121	6	2.0	310	1	KHSE_ECO57	Q8xa82 escherichia	194	1	BR53_MOUSE	O54798 mus musculu
122	6	2.0	310	1	KHSE_ECOLI	P00547 escherichia	195	1	YBFB_BACSU	O97967 ovis aries
123	6	2.0	311	1	MSAB_STRGC	Q9lam9 s peptide m	196	1	GSFP_VIECH	P37498 bacillus su
124	6	2.0	315	1	6DCS_SOYBN	P26690 glycine max	197	1	AROAF_THROA	P45780 vibrio chol
125	6	2.0	318	1	CH11_ORYSA	P24626 oryza sativ	198	1	Y588_BUCAI	Q97863 thermoplas
126	6	2.0	321	1	YECF_HAEIN	P44167 haemophilus	199	1	YM04_ARCFU	P57648 buchnera ap
127	6	2.0	321	1	SERB_ECOLI	P06862 escherichia	200	1	YNFM_ECOLI	Q28079 archaeoglob
128	6	2.0	325	1	IF32_HUMAN	Q13347 homo sapien	201	1	PROA_CLOAB	P43531 escherichia
129	6	2.0	325	1	IF32_MOUSE	Q9gzd9 mus musculu	202	1	K2C2_XENLA	Q97862 clostridium
130	6	2.0	325	1	WZB8_ECOLI	P35272 escherichia	203	1	RFAL_ECOLI	P04265 xenopus lae
131	6	2.0	325	1	WZB8_ECOLI	O33953 escherichia	204	1	YCG3_YEAST	P27243 escherichia
132	6	2.0	325	1	WZB8_SHIDY	P95730 shigella dy	205	1	YCG3_YEAST	P25591 saccharomyc
133	6	2.0	325	1	WZB8_SHIFL	P37792 shigella fl	206	1	ASGM_AERPE	Q9ybi2 aeropyrum p
134	6	2.0	326	1	WZB8_ECOLI	P76372 escherichia	207	1	K2C1_XENLA	P22051 tomato blac
135	6	2.0	331	1	HOLA_BUCAI	P57520 buchnera ap	208	1	PYRC_LISMO	P04266 xenopus lae
136	6	2.0	334	1	HEW2_HELPY	P56107 helicobacte	209	1	RMUC_RICPR	Q9y663 listeria mo
137	6	2.0	335	1	HEW2_HELPY	O9skd4 helicobacte	210	1	TRBI_RHISN	Q92cn3 rickettsia
138	6	2.0	336	1	RFAL_SALTY	P19817 salmonella	211	1	YHFT_ECOLI	P55406 rhizobium s
139	6	2.0	337	1	GO55_YEAST	P40107 saccharomyc	212	1	MURA_SYNY3	P45546 escherichia
140	6	2.0	337	1	MSAB_CAMFE	Q93kf3 c peptide m	213	1	Y281_HAEIN	Q55673 synechocyst
141	6	2.0	339	1	DCUP_HELPY	Q92lm8 helicobacte	214	1	IF2G_ENCCU	P44610 haemophilus
142	6	2.0	339	1	DCUP_HELPY	O52325 helicobacte	215	1	MUKF_ECOLI	Q96719 encephalito
143	6	2.0	339	1	HEM1_CHLPN	O927j1 chlamydia p	216	1	UGTC_CABEL	P36567 escherichia
144	6	2.0	339	1	STEA_HUMAN	Q9ueh8 homo sapien	217	1	YHJE_ECOLI	Q22181 caenorhabdi
145	6	2.0	341	1	BMPD_BORBU	Q44743 borrelia bu	218	1	Y314_MYGGE	P37643 escherichia
146	6	2.0	341	1	LP5L_RHIME	O54067 rhizobium m	219	1	YABL_HAEIN	Q9415 mycoplasma
147	6	2.0	342	1	P2YC_HUMAN	Q9h2k4 macaca fasc	220	1	G05A_DROME	P4936 haemophilus
148	6	2.0	342	1	P2YC_WACFA	Q95kc3 macaca fasc	221	1	FLDL_VIBPA	Q94975 vibrio para
149	6	2.0	342	1	SNXF_HUMAN	O9nr56 homo sapien	222	1	PH1_PSEPU	Q05181 pseudomonas
150	6	2.0	343	1	P2YC_RAT	Q9epx4 rattus norv	223	1	AK3_BACSU	P94417 bacillus su
151	6	2.0	349	1	PDXA_SYNY3	Q55982 synechocyst	224	1	DCUD_ECOLI	P45428 escherichia
152	6	2.0	349	1	PHOL_MYCLE	Q49751 mycobacteri	225	1	YC45_ODOSI	P45401 odontella s
153	6	2.0	351	1	HSF3_LCYPE	P41152 lycopersico	226	1	YNU6_YEAST	P40161 saccharomyc
154	6	2.0	351	1	VMAT_P13B	P06166 bovine para	227	1	ARLZ_SCHPO	P50514 schizosacch
155	6	2.0	353	1	MDHG_SOYBN	P37228 glycine max	228	1	ARLY_SCHPO	P40369 schizosacch
156	6	2.0	354	1	CKR5_MOUSE	P51682 mus musculu	229	1	MANA_EMENI	P29951 emericecia
157	6	2.0	354	1	CKR5_RAT	O08556 rattus norv	230	1	Y608_HAEIN	Q57486 haemophilus
158	6	2.0	356	1	TH14_CITSI	O23787 citrus sine	231	1	NHAC_BACFI	P27611 bacillus fi
159	6	2.0	357	1	DCUP_MYCLE	P46809 mycobacteri	232	1	VL2_FCPV1	Q02276 pygmy chimp
160	6	2.0	357	1	DCUP_MYCTU	O53231 mycobacteri	233	1	CRD1_CHICK	P02521 gallus gall
161	6	2.0	360	1	YAG3_MYCTU	O53411 mycobacteri	234	1	TRME_CYACA	Q9tlx6 cyanidium c
162	6	2.0	367	1	PROB_MYCLE	Q9cbz5 mycobacteri	235	1	CRD2_CHICK	P50583 gallus gall
163	6	2.0	369	1	RF2_CHLPN	P56906 chlamydia p	236	1	FUMC_ECOLI	Q05042 escherichia
164	6	2.0	373	1	CKR2_MOUSE	P51683 mus musculu	237	1	K2C7_HUMAN	P08729 homo sapien
165	6	2.0	373	1	CKR2_RAT	O55193 rattus norv	238	1	GLYC_YEAST	P37291 saccharomyc
166	6	2.0	374	1	CKR2_HUMAN	P41597 homo sapien	239	1	MUC1_HYLLA	Q29435 hylobates l
167	6	2.0	375	1	KOM8_YEAST	Q12222 saccharomyc	240	1	T23F_BACTI	Q02404 bacillus th
168	6	2.0	375	1	TR15_GIBZE	Q00909 gibberella	241	1	GATB_AQUAE	O66766 aquifex aeo
169	6	2.0	379	1	CYB_SORTU	O79464 sorex tundr	242	1	GLYC_MOUSE	P50431 mus musculu
170	6	2.0	379	1	MTS2_SHISO	P34879 shigella so	243	1	ATPB_CHLRE	P06541 chlamydomon
171	6	2.0	380	1	NTG2_YEAST	Q08214 saccharomyc	244	1	CG23_SCHPO	P10815 schizosacch
172	6	2.0	380	1	YRP4_SHVX	Q04583 shallot vir	245	1	SAHH_TRIVA	P51540 trichomonas
173	6	2.0	381	1	CYB_CARPL	P34866 carthariniu	246	1	CBP1_CANAL	P31225 candida alb
174	6	2.0	381	1	CYB_CARPO	P34867 carthariniu	247	1	GLG2_SYNY3	P72623 synechocyst
175	6	2.0	381	1	CYB_CARCU	P34868 galeocerdo	248	1	RBSA_HAEIN	P44735 haemophilus
176	6	2.0	381	1	CYB_NEGBR	P34872 negaprion b	249	1	TCMN_STRGA	P16559 streptomyce
177	6	2.0	381	1	CYB_PRIGL	P34873 prionace gl	250	1	CYSN_RHIME	P56893 rhizobium m
178	6	2.0	381	1	CYB_SPHLE	P34874 sphyrna lew	251	1	FEAB_ECOLI	P80668 escherichia
179	6	2.0	381	1	CYB_SPHTT	P34876 sphyrna tib	252	1	RBSA_ECOLI	P04983 escherichia

253	6	2.0	502	1	K2C8_XENLA	P08776 xenopus lae	326	6	2.0	604	1	YFIC_BACSU	P54719 bacillus su
254	6	2.0	502	1	K2M3_SHEEP	P25691 ovis aries	327	6	2.0	605	1	YHC8_YEAST	P38739 saccharomyc
255	6	2.0	502	1	UAP1_ARATH	O64765 arabidopsis	328	6	2.0	607	1	HEPA_ANASP	P22638 anabaena sp
256	6	2.0	502	1	YQCL_ECOLI	O46901 escherichia	329	6	2.0	612	1	AMVG_ASPOR	P36914 aspergillus
257	6	2.0	503	1	MTBR_BPH1	P09915 bacterioph	330	6	2.0	612	1	GIDA_MYCPN	P75221 mycoplasma
258	6	2.0	504	1	ATPA_WHEAT	P12112 triticum ae	331	6	2.0	614	1	CYSN_MYCTU	Q10600 m cyan/cyoc
259	6	2.0	505	1	NUOM_BUCAL	P57263 buchnera ap	332	6	2.0	616	1	ORC3_YEAST	P54790 saccharomyc
260	6	2.0	506	1	ATPA_EUGR	P30392 euglena gra	333	6	2.0	620	1	NODQ_AZOBR	P28604 a nodq bifu
261	6	2.0	507	1	ATPA_MAIZE	P05022 zea mays (m	334	6	2.0	623	1	HPC2_YEAST	Q01448 saccharomyc
262	6	2.0	507	1	ATPA_ORYSA	P12084 oryza sativ	335	6	2.0	627	1	K2C1_MOUSE	Q04104 mus musculu
263	6	2.0	507	1	CBS_YEAST	P32582 saccharomyc	336	6	2.0	629	1	PAB2_ARATH	P42731 arabidopsis
264	6	2.0	508	1	CP77_HORSE	Q95328 equus caball	337	6	2.0	630	1	TNPB_STRAU	P06637 staphylococ
265	6	2.0	509	1	YRW4_CABEL	Q10051 caenorhabdi	338	6	2.0	630	1	TNPE_STRAU	P37375 staphylococ
266	6	2.0	510	1	G6PD_ASPNG	P48826 aspergillus	339	6	2.0	631	1	OE66_NPVL5	P89518 leucania se
267	6	2.0	510	1	NIFB_FRAAL	P46044 frankia aln	340	6	2.0	633	1	NODQ_RHISB	Q07309 r nodq bifu
268	6	2.0	510	1	PUR1_YEAST	P40406 saccharomyc	341	6	2.0	633	1	Y561_HAEIN	P44016 haemophilus
269	6	2.0	511	1	G6PD_EMENI	P41764 emeritocella	342	6	2.0	635	1	DALA_YEAST	Q04895 saccharomyc
270	6	2.0	512	1	CADC_ECOLI	P31890 escherichia	343	6	2.0	638	1	K220_HUMAN	Q01546 homo sapien
271	6	2.0	512	1	GLGT_VICPA	P52417 vicia faba	344	6	2.0	639	1	GLGB_BUTFI	P30539 butyrivibri
272	6	2.0	512	1	K2C5_XENLA	P16878 xenopus lae	345	6	2.0	639	1	GYRB_BORBU	P33769 borrelia bu
273	6	2.0	515	1	GLCM_MOUSE	P17439 mus musculu	346	6	2.0	639	1	Y119_MYCTU	Q50614 mycobacteri
274	6	2.0	515	1	LEU1_BACST	Q8r185 bacillus st	347	6	2.0	640	1	AMVG_ASPNG	P04064 aspergillus
275	6	2.0	517	1	LEGA_PEA	P02857 pisum sativ	348	6	2.0	641	1	NODQ_RHIME	P13442 r nodq bifu
276	6	2.0	520	1	LAC1_CORHI	Q02497 coriolus hi	349	6	2.0	645	1	K22E_HUMAN	P35908 homo sapien
277	6	2.0	520	1	LAC1_TRAVI	Q99044 trametes vi	350	6	2.0	648	1	AMYA_PYRFU	P49067 pyrococcus
278	6	2.0	524	1	LEG2_PEA	P15838 pisum sativ	351	6	2.0	648	1	GRT1_SCHPO	Q9c459 schizosacch
279	6	2.0	524	1	K2C4_MOUSE	P07744 mus musculu	352	6	2.0	650	1	MTLR_STRMU	Q02425 streptococ
280	6	2.0	525	1	ACU8_NEUCR	P51937 neurospora	353	6	2.0	651	1	Y1X1_SCHPO	Q9p713 schizosacch
281	6	2.0	525	1	PUR9_BUCAL	P57143 b bifunctio	354	6	2.0	653	1	MALQ_PYRKO	Q32450 pyrococcus
282	6	2.0	528	1	VNH7_YEAST	P53940 saccharomyc	355	6	2.0	655	1	CDGT_KLEPN	P08704 klebsiella
283	6	2.0	529	1	TYRO_CHICK	P55024 gallus gall	356	6	2.0	656	1	YID7_YEAST	P40534 saccharomyc
284	6	2.0	529	1	TYRO_HUMAN	P14679 homo sapien	357	6	2.0	662	1	MTRI_HUMAN	Q13613 homo sapien
285	6	2.0	531	1	Y4TO_RHISN	P55669 rhizobium s	358	6	2.0	671	1	CHEA_BACSU	P29072 bacillus su
286	6	2.0	534	1	K2C4_HUMAN	P19013 homo sapien	359	6	2.0	674	1	YB66_YEAST	P38315 saccharomyc
287	6	2.0	535	1	YHIL_ECOLI	P37629 escherichia	360	6	2.0	676	1	KPCD_HUMAN	Q05655 homo sapien
288	6	2.0	536	1	GVD1_HALNI	P32043 halobacteri	361	6	2.0	680	1	YH2X_SCHPO	Q74343 schizosacch
289	6	2.0	540	1	BB61_RABIT	Q05004 oryctolagus	362	6	2.0	685	1	RNF6_HUMAN	Q9252 homo sapien
290	6	2.0	540	1	TYRO_ORYLA	P55025 oryzias lat	363	6	2.0	689	1	SP11_HUMAN	Q9hb58 homo sapien
291	6	2.0	543	1	YJF8_YEAST	P47041 saccharomyc	364	6	2.0	697	1	SSRP_CABEL	P41848 caenorhabdi
292	6	2.0	544	1	PYRG_ECOLI	P08398 escherichia	365	6	2.0	702	1	FTSH_MYCGE	P47695 mycoplasma
293	6	2.0	552	1	K2C6_MOUSE	P50446 mus musculu	366	6	2.0	706	1	KPCT_HUMAN	Q04759 homo sapien
294	6	2.0	555	1	C166_CARAU	Q90304 carassius a	367	6	2.0	707	1	KPCT_MOUSE	Q02111 mus musculu
295	6	2.0	556	1	PUR1_MYCLE	Q50028 mycobacteri	368	6	2.0	709	1	FTSH_MYCPN	P75120 mycoplasma
296	6	2.0	558	1	CSG_METJA	Q58232 methanococ	369	6	2.0	711	1	PNP_ECOLI	P05055 escherichia
297	6	2.0	560	1	YB70_VACCV	P24769 vaccinia vi	370	6	2.0	729	1	SYG_ARATH	Q23627 arabidopsis
298	6	2.0	560	1	YB70_YEAST	P38318 saccharomyc	371	6	2.0	733	1	PTK_AC1JO	O52788 acinetobact
299	6	2.0	561	1	O16G_BACSU	O06994 bacillus su	372	6	2.0	733	1	YACK_RHIME	Q9x447 rhizobium m
300	6	2.0	562	1	YX13_CABEL	Q11124 caenorhabdi	373	6	2.0	734	1	GL3B_DICDI	P34116 dictyostell
301	6	2.0	563	1	K2CA_HUMAN	P02538 homo sapien	374	6	2.0	734	1	GLGB_AGRTO	P52979 agrobacteri
302	6	2.0	563	1	K2CB_HUMAN	P04259 homo sapien	375	6	2.0	734	1	GLGB_AGRTO	P28969 equine herp
303	6	2.0	563	1	K2CC_HUMAN	P48666 homo sapien	376	6	2.0	738	1	AB10_HUMAN	Q9nrk6 homo sapien
304	6	2.0	563	1	K2CE_HUMAN	P48668 homo sapien	377	6	2.0	740	1	PEC1_PIG	Q95242 sus acrofa
305	6	2.0	563	1	K2CF_HUMAN	P48669 homo sapien	378	6	2.0	741	1	BSG2_DROME	P11929 drosophila
306	6	2.0	572	1	LMD1_HUMAN	P29536 homo sapien	379	6	2.0	743	1	KPC2_APLCA	Q16975 aplysia cal
307	6	2.0	578	1	PTFB_RHOCA	P33387 rhodobacter	380	6	2.0	748	1	MEPA_RAT	Q64230 rattus norv
308	6	2.0	578	1	VAC8_YEAST	P33968 saccharomyc	381	6	2.0	758	1	SC18_YEAST	P18759 saccharomyc
309	6	2.0	579	1	UVRC_METTH	O26541 methanobact	382	6	2.0	759	1	YDEP_ECOLI	P77561 escherichia
310	6	2.0	583	1	EXP8_STRPN	P35598 streptococ	383	6	2.0	760	1	METE_MYCLE	O05564 mycobacteri
311	6	2.0	589	1	ACM3_MOUSE	Q9er23 mus musculu	384	6	2.0	764	1	TSHE_MOUSE	P47750 mus musculu
312	6	2.0	589	1	ACM3_RAT	P08483 rattus norv	385	6	2.0	764	1	TSHE_MOUSE	P47750 mus musculu
313	6	2.0	589	1	YAN6_SCHPO	Q10072 schizosacch	386	6	2.0	772	1	ANNU_SCHAM	P21463 schistocerc
314	6	2.0	590	1	ACM3_BOVIN	P41984 bos taurus	387	6	2.0	774	1	YMA6_YEAST	Q04233 saccharomyc
315	6	2.0	590	1	ACM3_GORGO	Q9n2a3 gorilla gor	388	6	2.0	776	1	CG2F_MOUSE	P51944 mus musculu
316	6	2.0	590	1	ACM3_HUMAN	P20309 homo sapien	389	6	2.0	780	1	NH48_CABEL	Q94407 caenorhabdi
317	6	2.0	590	1	ACM3_PANTR	Q9n2a4 pan troslod	390	6	2.0	786	1	CG2F_HUMAN	P41002 homo sapien
318	6	2.0	590	1	ACM3_PIG	P11483 sus scrofa	391	6	2.0	790	1	QUIA_XANCU	Q9xd78 xanthomonas
319	6	2.0	590	1	ACM3_PONPY	Q9n2a2 pongo pygma	392	6	2.0	798	1	HMPB_DROME	P31264 drosophila
320	6	2.0	590	1	K2C5_HUMAN	P13647 homo sapien	393	6	2.0	798	1	RP94_FOWPV	Q9j589 fowlpox vir
321	6	2.0	599	1	REF2_DROME	P14199 drosophila	394	6	2.0	810	1	KIJ5_YEAST	P40494 saccharomyc
322	6	2.0	599	1	REF2P_DROSI	Q24629 drosophila	395	6	2.0	810	1	PFLF_ECOLI	P75753 escherichia
323	6	2.0	600	1	KU70_RHIAP	Q26228 rhipicephal	396	6	2.0	810	1	SYFB_SYNY3	P74236 synechocyst
324	6	2.0	602	1	LEPA_HELPFU	Q95122 helicobacte	397	6	2.0	811	1	GAG_FOAMV	P14349 human spuma
325	6	2.0	604	1	LEPA_HELPFU	Q9zm93 helicobacte	398	6	2.0	815	1	AKH_HAEIN	P44505 haemophilus

399	6	2.0	827	1	GYRA_HELPY	P48370 helicobacte	472	6	2.0	1335	1	DP3A_DEIRA	Q9rx08 deinococcu
400	6	2.0	856	1	HMDH_BIAGE	P54960 blattella g	473	6	2.0	1388	1	HRP3_SCHPO	O14139 schizosacch
401	6	2.0	856	1	POLG_PVYC	P22601 potato viru	474	6	2.0	1403	1	CIC_DROME	Q9ulh0 drosophila
402	6	2.0	856	1	POLG_PVYO	P22602 potato viru	475	6	2.0	1411	1	Y197_DROME	Q9vkl3 drosophila
403	6	2.0	859	1	ENV_ETAV2	P22428 equine infe	476	6	2.0	1443	1	NYJ2_HUMAN	O15056 homo sapien
404	6	2.0	860	1	ENV_ETAV5	P22430 equine infe	477	6	2.0	1460	1	N159_YEAST	P40477 saccharomyc
405	6	2.0	877	1	YEST_ECOLI	P76272 escherichia	478	6	2.0	1463	1	PA2R_BOVIN	P49259 bos taurus
406	6	2.0	881	1	GLND_VIBCH	Q9kpv0 vibrio chol	479	6	2.0	1472	1	A2MG_RAT	P06238 rattus norv
407	6	2.0	886	1	VGP3_EBVA8	Q07284 epstein-bar	480	6	2.0	1567	1	FMN2_MOUSE	Q9j104 mus musculu
408	6	2.0	899	1	PMAL_KLULA	P49380 kluyveromyc	481	6	2.0	1630	1	ESPI_YEAST	Q03018 saccharomyc
409	6	2.0	901	1	GREB_BORBU	O51157 borrelia bu	482	6	2.0	1668	1	DPOL_THERY	Q9hh05 thermococcu
410	6	2.0	906	1	ACON_BRAJA	P70920 bradyrhizob	483	6	2.0	1690	1	RPOC_TBEMA	P36252 thermotoga
411	6	2.0	907	1	GACS_PSESY	P48027 pseudomonas	484	6	2.0	1705	1	CYAA_BORBR	Q57506 bordetella
412	6	2.0	907	1	VGP3_EBV	P03200 epstein-bar	485	6	2.0	1706	1	CYAA_BORPE	P15318 bordetella
413	6	2.0	916	1	SECA_TREPA	O83394 treponema p	486	6	2.0	1755	1	YJZ7_YEAST	P47098 saccharomyc
414	6	2.0	918	1	PMAL_YEAST	P05030 saccharomyc	487	6	2.0	1755	1	YJZ9_YEAST	P47100 saccharomyc
415	6	2.0	918	1	QALS_NEUCR	P11637 neurospora	488	6	2.0	1803	1	YJL3_YEAST	P47024 saccharomyc
416	6	2.0	919	1	PWP2_HUMAN	Q15269 homo sapien	489	6	2.0	1872	1	T2D1_HUMAN	P21675 homo sapien
417	6	2.0	920	1	PMAL_ZYGRO	P24545 zygosacchar	490	6	2.0	1887	1	FAS2_YEAST	P19097 s fatty aci
418	6	2.0	920	1	SYI_HELPY	Q92j11 helicobacte	491	6	2.0	2137	1	SPCB_HUMAN	P11277 homo sapien
419	6	2.0	920	1	SYI_HELPY	P56456 helicobacte	492	6	2.0	2165	1	RRPL_HRSVA	P28887 human respi
420	6	2.0	932	1	SSA1_PASHA	P31631 pasteurella	493	6	2.0	2208	1	POLN_MANCV	Q69014 manchester
421	6	2.0	933	1	MSH2_RAT	P54275 rattus norv	494	6	2.0	2226	1	POLG_HPAV2	P26580 hepatitis a
422	6	2.0	934	1	MSH2_HUMAN	P43246 homo sapien	495	6	2.0	2226	1	POLG_HPAV4	P26581 hepatitis a
423	6	2.0	935	1	MSH2_MOUSE	P43247 mus musculu	496	6	2.0	2226	1	POLG_HPAV8	P26582 hepatitis a
424	6	2.0	937	1	SYL_METHH	O27552 methanobact	497	6	2.0	2227	1	POLG_HPAVL	P08617 hepatitis a
425	6	2.0	943	1	UVRA_HAEIN	P44410 haemophilus	498	6	2.0	2227	1	POLG_HPAVL	P06441 hepatitis a
426	6	2.0	947	1	PMAL_YEAST	P19657 saccharomyc	499	6	2.0	2227	1	POLG_HPAVM	P13901 hepatitis a
427	6	2.0	948	1	UVRA_MYCPN	P75176 mycoplasma	500	6	2.0	2230	1	GOG4_HUMAN	Q13439 homo sapien
428	6	2.0	952	1	UVRA_MYCGE	P47660 mycoplasma	501	6	2.0	2230	1	POLG_HPAVS	P14553 simian hepa
429	6	2.0	953	1	YCQ7_YEAST	P25618 saccharomyc	502	6	2.0	2469	1	TEGU_HSVSA	Q1056 herpesvirus
430	6	2.0	957	1	UVRA_BACHD	Q9k6y0 bacillus ha	503	6	2.0	2493	1	CYAA_USTMA	P49606 ustilago ma
431	6	2.0	960	1	CSE1_YEAST	P33307 saccharomyc	504	6	2.0	2670	1	YAO5_SCHPO	Q10105 schizosacch
432	6	2.0	960	1	UVRA_TREPA	O83527 treponema p	505	6	2.0	2769	1	THYG_BOVIN	P01267 bos taurus
433	6	2.0	970	1	UVRA_SVNY3	P73412 synechocyst	506	6	2.0	2875	1	RRPL_TSWV1	P28976 tomato spot
434	6	2.0	988	1	M172_MOUSE	P97432 mus musculu	507	6	2.0	3010	1	POLG_HCVUT	Q00269 h genome po
435	6	2.0	993	1	SMB2_HUMAN	P39935 homo sapien	508	6	2.0	3063	1	POLG_PVTN	P18247 p genome po
436	6	2.0	999	1	HGP4_HAEIN	Q57408 haemophilus	509	6	2.0	3305	1	APLP_MANSE	Q25490 manduca sex
437	6	2.0	1005	1	DPOL_VARV	P33793 variola vir	510	6	2.0	3411	1	POLG_YEFV1	P03314 y genome po
438	6	2.0	1006	1	DPOL_VACCC	P20509 vaccinia vi	511	6	2.0	3411	1	POLG_YEFV2	P19901 y genome po
439	6	2.0	1006	1	DPOL_VACCV	Q06856 vaccinia vi	512	6	2.0	3712	1	ACVS_CEPAC	P25464 cephalospor
440	6	2.0	1039	1	YR71_CAEEL	Q09564 caenorhabdi	513	6	2.0	4499	1	DYHA_CHLRE	Q39610 chlamydomon
441	6	2.0	1040	1	RIK1_SCHPO	Q10426 schizosacch	514	6	2.0	4563	1	APB_HUMAN	P04114 homo sapien
442	6	2.0	1045	1	PRTS_SERMA	P09489 serratia ma	515	6	2.0	4687	1	PLE1_RAT	P30427 rattus norv
443	6	2.0	1045	1	PRTT_SERMA	P29805 serratia ma	516	6	2.0	4687	1	HTS1_COCCA	Q1886 cochllobolu
444	6	2.0	1053	1	SLPM_BACBR	P06546 bacillus br	517	5	1.7	21	MDH_EURCE	P80537 burkholderi	
445	6	2.0	1062	1	CC7_SCHPO	P41892 schizosacch	518	5	1.7	21	MDH_PSEIN	P80538 pseudomonas	
446	6	2.0	1064	1	CYA4_RAT	P26770 rattus norv	519	5	1.7	29	MDH_BURPS	P80536 burkholderi	
447	6	2.0	1077	1	XVNY_CLOTM	P51584 clostridium	520	5	1.7	33	YC12_NEPOL	Q9tky6 nephroselmi	
448	6	2.0	1078	1	S24A_HUMAN	O95486 homo sapien	521	5	1.7	36	RL6_HALCU	P05968 halobacteri	
449	6	2.0	1094	1	A3B1_HUMAN	O00203 homo sapien	522	5	1.7	40	GP55_BPSP1	Q48409 bacterioph	
450	6	2.0	1099	1	NKC2_RABIT	P55015 oryctolagus	523	5	1.7	50	PND1_ECOLI	P1902 escherichia	
451	6	2.0	1105	1	A3B1_MOUSE	Q921t1 mus musculu	524	5	1.7	50	PND2_ECOLI	P16477 escherichia	
452	6	2.0	1113	1	HDAS_MOUSE	Q922v6 mus musculu	525	5	1.7	50	S10D_CHICK	P51964 gallus gall	
453	6	2.0	1122	1	HDAS_HUMAN	Q9ucl6 homo sapien	526	5	1.7	50	ZNT4_BOVIN	Q8ttf3 bos taurus	
454	6	2.0	1148	1	YKQ0_CAEEL	P34305 caenorhabdi	527	5	1.7	52	NAHQ_PSEPU	Q51945 pseudomonas	
455	6	2.0	1176	1	VPS8_YEAST	P39702 saccharomyc	528	5	1.7	53	LECA_VICCR	P02868 vicia cracc	
456	6	2.0	1185	1	MAPX_DROME	P23226 drosophila	529	5	1.7	58	VG58_BPPFP3	P03629 bacterioph	
457	6	2.0	1232	1	B3A3_HUMAN	P48751 homo sapien	530	5	1.7	60	HMEN_LAMPL	P31534 lampetra pl	
458	6	2.0	1238	1	BVGS_BORBR	P26762 bordetella	531	5	1.7	62	PETL_PINTH	P52805 pinus thunb	
459	6	2.0	1238	1	BVGS_BORPA	P40330 bordetella	532	5	1.7	62	RM32_RECAM	O21281 reclinomona	
460	6	2.0	1238	1	BVGS_BORPE	P16575 bordetella	533	5	1.7	63	JHEB_TRINI	P30810 trichoplusi	
461	6	2.0	1255	1	MUC1_HUMAN	P15941 h mucin 1 p	534	5	1.7	64	EAP_BOVIN	O02775 bos taurus	
462	6	2.0	1256	1	MRP_STRSU	P32653 streptococc	535	5	1.7	64	V07K_CVB	P37990 chrysanthem	
463	6	2.0	1305	1	CYA9_XENLA	Q98999 xenopus lae	536	5	1.7	64	Y592_HAEIN	P44021 haemophilus	
464	6	2.0	1309	1	ST16_SCHPO	Q09743 schizosacch	537	5	1.7	66	V07K_WCMVM	P09501 white clove	
465	6	2.0	1311	1	FMR2_HUMAN	P51816 homo sapien	538	5	1.7	67	ATP8_CANPA	Q92263 canis famil	
466	6	2.0	1313	1	VGLM_PTPV	P03517 punta toro	539	5	1.7	67	TATE_SALTY	P57050 salmonella	
467	6	2.0	1328	1	EXO2_SCHPO	P40383 schizosacch	540	5	1.7	67	YJIX_ECOLI	P39395 escherichia	
468	6	2.0	1328	1	YMD9_YEAST	Q03434 saccharomyc	541	5	1.7	67	YL32_HALMA	P20528 haloarcula	
469	6	2.0	1328	1	YME4_YEAST	Q04711 saccharomyc	542	5	1.7	67	YVAS_VACCV	P20572 vaccinia vi	
470	6	2.0	1328	1	YMT5_YEAST	Q04214 saccharomyc	543	5	1.7	68	LPPL_MOUSE	P97400 mus musculu	
471	6	2.0	1328	1	YMU0_YEAST	Q04670 saccharomyc	544	5	1.7	68	RR18_CYACA	Q9tm39 cyanidium c	

545	1.7	69	1	Y132_TREPA	083168	treponema p	618	5	1.7	96	1	CH10_HOLOB	P94819	holoepora o
546	1.7	70	1	Y178_THETH	P43891	thermus the	619	5	1.7	96	1	IF1C_TOBAC	P12136	nicotiana t
547	1.7	71	1	Y178_THETH	Q82yf4	pyrobaculum	620	5	1.7	96	1	LYOX_BOVIN	P33072	bos laurus
548	1.7	71	1	Y178_THETH	P09046	porcine tra	621	5	1.7	96	1	SV01_HUMAN	P23262	homo sapien
549	1.7	71	1	Y178_THETH	P05650	bacillus su	622	5	1.7	96	1	BORD_ECOLI	P77330	escherichia
550	1.7	71	1	Y178_THETH	P17359	vaccinia vi	623	5	1.7	96	1	BORD_ECOLI	P26814	bacterioph
551	1.7	72	1	Y178_THETH	P05457	bradyrhizob	624	5	1.7	96	1	FLGM_ECOLI	P43532	escherichia
552	1.7	72	1	Y178_THETH	P35138	bacillus li	625	5	1.7	96	1	FLGM_ECOLI	P26477	salmonella
553	1.7	73	1	Y178_THETH	P03617	bacterioph	626	5	1.7	96	1	GAC2_CLOAB	Q97ex7	clostridium
554	1.7	73	1	Y178_THETH	083550	treponema p	627	5	1.7	96	1	RL31_MYCGE	P47499	mycoplasma
555	1.7	74	1	Y178_THETH	Q1652	anser caeru	628	5	1.7	96	1	Y404_ANASP	Q8vzg3	anabaena ap
556	1.7	75	1	Y178_THETH	Q9fsl0	asparagus o	629	5	1.7	96	1	ACYM_HORSE	P00818	equus cabal
557	1.7	75	1	Y178_THETH	Q9fsl9	cyandium c	630	5	1.7	96	1	FIXX_RHILT	P008710	rhizobium l
558	1.7	76	1	Y178_THETH	P37085	onocotnes	631	5	1.7	96	1	GVJ2_HALN1	P33956	halobacteri
559	1.7	76	1	Y178_THETH	P20522	vaccinia vi	632	5	1.7	96	1	PLAS_ENTPR	P7465	enteromorph
560	1.7	77	1	Y178_THETH	P1065	escherichia	633	5	1.7	96	1	PLAS_ULVPE	P13133	ulva araeak
561	1.7	77	1	Y178_THETH	P20526	vaccinia vi	634	5	1.7	96	1	PLAS_ULVPE	P56274	ulva pertus
562	1.7	78	1	Y178_THETH	Q53972	shigella dy	635	5	1.7	96	1	Y097_ARCFU	Q30139	archaeoglob
563	1.7	78	1	Y178_THETH	Q14781	homo sapien	636	5	1.7	96	1	APC3_HUMAN	P2656	homo sapien
564	1.7	78	1	Y178_THETH	Q06202	carcinus ma	637	5	1.7	96	1	CML_SYN7	Q03512	synechococc
565	1.7	78	1	Y178_THETH	P51727	bacterioph	638	5	1.7	96	1	DHGP_ASPNG	P5804	aspergillus
566	1.7	78	1	Y178_THETH	Q92686	homo sapien	639	5	1.7	96	1	ETBR_MACFA	Q38468	macaca faec
567	1.7	78	1	Y178_THETH	P97816	mus musculus	640	5	1.7	96	1	IFHA_XANCP	Q56767	xanthomonas
568	1.7	78	1	Y178_THETH	P02634	rattus norv	641	5	1.7	96	1	PLAS_LACSA	P00290	lactuca sat
569	1.7	79	1	Y178_THETH	Q92445	paracoccidi	642	5	1.7	96	1	PLAS_PHAUV	P00287	phaseolus v
570	1.7	80	1	Y178_THETH	P58217	escherichia	643	5	1.7	96	1	PLAS_RUNOB	P00298	rumex obtus
571	1.7	80	1	Y178_THETH	O10311	orygia pseu	644	5	1.7	96	1	PLAS_SANMI	P00291	sambucus ni
572	1.7	81	1	Y178_THETH	P45309	haemophilus	645	5	1.7	96	1	PLAS_SOLCR	P00297	solanum cri
573	1.7	81	1	Y178_THETH	P42970	bacillus su	646	5	1.7	96	1	PLAS_SOLTU	P00296	solanum tub
574	1.7	82	1	Y178_THETH	Q9eyf1	escherichia	647	5	1.7	96	1	PLAT_TOBAC	P35477	nicotiana t
575	1.7	82	1	Y178_THETH	P58218	escherichia	648	5	1.7	96	1	RL23_BUCAT	P57589	buchnera ap
576	1.7	82	1	Y178_THETH	Q8z495	salmonella	649	5	1.7	96	1	URE3_YEREN	P31496	yersinia en
577	1.7	82	1	Y178_THETH	P56035	salmonella	650	5	1.7	96	1	URE3_YERPE	P52315	yersinia pe
578	1.7	82	1	Y178_THETH	P18643	salmonella	651	5	1.7	96	1	ELAC_TRIVU	Q29143	trichosurus
579	1.7	83	1	Y178_THETH	Q06079	shigella fl	652	5	1.7	96	1	Y57_PAPVD	P03131	deer papill
580	1.7	83	1	Y178_THETH	Q8x713	rhizobium m	653	5	1.7	96	1	YALJ_ICNV	Q8588	indian cass
581	1.7	84	1	Y178_THETH	Q33602	staphylococ	654	5	1.7	96	1	YC4_TYLCA	P36283	tomato yell
582	1.7	84	1	Y178_THETH	P27385	bacterioph	655	5	1.7	96	1	CHLB_SELMO	P37856	selaginella
583	1.7	84	1	Y178_THETH	P20212	sulfolobus	656	5	1.7	96	1	YA40_MYCPN	P75074	mycoplasma
584	1.7	85	1	Y178_THETH	P00111	porphyra te	657	5	1.7	96	1	MGP_MOUSE	P19788	mus musculus
585	1.7	85	1	Y178_THETH	Q9ia08	dicentrarch	658	5	1.7	96	1	Y327_METJA	Q57773	methanococc
586	1.7	85	1	Y178_THETH	P37044	haplochromi	659	5	1.7	96	1	YGE1_YEAST	P33182	saccharomyc
587	1.7	85	1	Y178_THETH	O73811	morone saxa	660	5	1.7	96	1	LV4B_HUMAN	P01716	homo sapien
588	1.7	85	1	Y178_THETH	P51925	sparus aura	661	5	1.7	96	1	LV4D_HUMAN	P01718	homo sapien
589	1.7	85	1	Y178_THETH	P49394	borrelia bu	662	5	1.7	96	1	LV4E_HUMAN	P06889	homo sapien
590	1.7	85	1	Y178_THETH	Q57960	methanococc	663	5	1.7	96	1	YQCC_HABIN	Q57152	haemophilus
591	1.7	86	1	Y178_THETH	O42241	oncorhynch	664	5	1.7	96	1	ATPK_DROME	Q8w141	drosophila
592	1.7	86	1	Y178_THETH	O83440	treponema p	665	5	1.7	96	1	LV4C_HUMAN	P01717	homo sapien
593	1.7	86	1	Y178_THETH	P18331	escherichia	666	5	1.7	96	1	VSMP_IBVU5	P30248	avian infec
594	1.7	87	1	Y178_THETH	O28908	archaeoglob	667	5	1.7	96	1	YGIY_YEAST	P53232	saccharomyc
595	1.7	87	1	Y178_THETH	P18585	chlamydia t	668	5	1.7	96	1	KV05_RABIT	P01686	oryctolagus
596	1.7	88	1	Y178_THETH	P21355	chlamydia t	669	5	1.7	96	1	KV10_HUMAN	P01609	homo sapien
597	1.7	88	1	Y178_THETH	Q97jko	clostridium	670	5	1.7	96	1	KV1V_HUMAN	P04430	homo sapien
598	1.7	89	1	Y178_THETH	Q1507	caenorhabdi	671	5	1.7	96	1	LV5A_HUMAN	P01719	homo sapien
599	1.7	89	1	Y178_THETH	Q9et21	oryza sativ	672	5	1.7	96	1	Y04B_BPT4	P07082	bacterioph
600	1.7	89	1	Y178_THETH	Q84850	chlamydia t	673	5	1.7	96	1	YC54_PORPU	P51204	porphyra pu
601	1.7	89	1	Y178_THETH	P42295	bacillus su	674	5	1.7	96	1	FEL2_FELCA	P30440	felis silve
602	1.7	90	1	Y178_THETH	P46217	sulfolobus	675	5	1.7	96	1	JHEA_TRINI	P30809	trichoplusi
603	1.7	90	1	Y178_THETH	P57596	buchnera ap	676	5	1.7	96	1	NO16_SOYBN	P23233	glycine max
604	1.7	90	1	Y178_THETH	P18633	onocotnes	677	5	1.7	96	1	PRVA_FELCA	P80079	felis silve
605	1.7	91	1	Y178_THETH	Q2107	acinetobact	678	5	1.7	96	1	PRVA_RABIT	P02624	oryctolagus
606	1.7	91	1	Y178_THETH	P04131	pseudomonas	679	5	1.7	96	1	SU11_SCHPO	P79060	schizosacch
607	1.7	91	1	Y178_THETH	P13113	serratia ma	680	5	1.7	96	1	COX4_BACSU	P24013	bacillus su
608	1.7	91	1	Y178_THETH	P04129	shigella fl	681	5	1.7	96	1	CYC6_PORPU	P51200	porphyra pu
609	1.7	91	1	Y178_THETH	Q9zda4	rickettsia	682	5	1.7	96	1	HIS2_CLOAB	Q78khs	clostridium
610	1.7	92	1	Y178_THETH	Q9z233	mus musculus	683	5	1.7	96	1	PRVA_RANCA	P18087	rana catesb
611	1.7	92	1	Y178_THETH	Q08589	rattus norv	684	5	1.7	96	1	RLAI_CLAHE	P50344	cladosporiu
612	1.7	92	1	Y178_THETH	Q57574	methanococc	685	5	1.7	96	1	RR18_PEA	P49169	pisum sativ
613	1.7	93	1	Y178_THETH	Q58490	methanococc	686	5	1.7	96	1	YK21_CABEL	P34328	caenorhabdi
614	1.7	93	1	Y178_THETH	Q26234	methanobact	687	5	1.7	96	1	YK21_CABEL	P34328	caenorhabdi
615	1.7	93	1	Y178_THETH	Q50799	methanobact	688	5	1.7	96	1	IR06_HCNVA	P96468	streptococc
616	1.7	93	1	Y178_THETH	P58251	clostridium	689	5	1.7	96	1	KV3A_MOUSE	P16804	human cytom
617	1.7	93	1	Y178_THETH	Q9y3y2	homo sapien	690	5	1.7	96	1	KV3C_MOUSE	P01656	mus musculus





983 5 1.7 160 1 IAGB\_SALTY P43017 salmonella  
 984 5 1.7 160 1 M02\_CRIGR P58755 cricetus  
 985 5 1.7 160 1 RISB\_MYCTU P71685 mycobacteri  
 986 5 1.7 160 1 THY1\_CHICK Q07212 gallus gall  
 987 5 1.7 160 1 VLPC\_MYCHR P29230 mycoplasma  
 988 5 1.7 160 1 IBP\_BUCAL Q92616 buchnera ap  
 989 5 1.7 160 1 PHAC\_PORPU P51195 porphyra pu  
 990 5 1.7 160 1 PTGA\_BUCAL Q9wx17 buchnera ap  
 991 5 1.7 160 1 YAN8\_YEAST P39564 saccharomyc  
 992 5 1.7 160 1 AROK\_LACLC P43906 lactococcus  
 993 5 1.7 160 1 COAD\_XYLEFA Q9pep8 xylella fas  
 994 5 1.7 160 1 CYPH\_CANAL P22011 candida alb  
 995 5 1.7 160 1 HSBX\_EAT P97541 rattus norv  
 996 5 1.7 160 1 MENG\_PSEAB Q912w7 pseudomonas  
 997 5 1.7 160 1 RHO\_BUCAP O51891 buchnera ap  
 998 5 1.7 160 1 RRPO\_LVX P27327 lily virus  
 999 5 1.7 160 1 SODM\_CORDI P42821 corynebacte  
 1000 5 1.7 160 1 VA49\_VACCC P21068 vaccinia vi

## ALIGNMENTS

RESULT 1  
 RL10\_NEIMA  
 ID RL10\_NEIMA STANDARD; PRT; 166 AA.  
 AC Q9QCP7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 508 ribosomal protein L10.  
 GN RPLJ OR NMA0144 OR NMB0130.  
 OS Neisseria meningitidis (serogroup A), and  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=65699, 491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=22491 / Serogroup A / Serotype 4A;  
 RX MEDLINE=20222556; PubMed=10761919;  
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
 RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,  
 RA Jagals K., Leather S., Meule S., Mungall K., Quail M.A.,  
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
 RA Whitehead S., Spratt B.G., Barrall B.G.;  
 RT "Complete DNA sequence of a serogroup A strain of Neisseria  
 meningitidis 22491."  
 RL Nature 404:502-506(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WC58 / Serogroup B;  
 RX MEDLINE=2017555; PubMed=10710307;  
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
 RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,  
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
 RA Mason T., Ciecko A., Parksey D.S., Blair E., Citterone H., Clark E.B.,  
 RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,  
 RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,  
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;  
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
 MC58."  
 RL Science 287:1809-1815(2000).  
 RN [3]

CC -!- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AL162752; CAB83459.1; -;  
 DR EMBL; AE002371; AAF40589.1; -;  
 DR TIGR; NMB0130; -;  
 DR InterPro; IPR001790; Ribosomal\_L10.  
 DR InterPro; IPR002363; Ribosomal\_L10eub.  
 DR Pfam; PF00466; Ribosomal\_L10; 1.  
 DR PROSITE; PS01109; RIBOSOMAL\_L10; 1.  
 KW Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 166 AA; 17594 MW; E90B1423F65317EE CRC64;  
 Query Match 2.7%; Score 8; DB 1; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 262 VAEIASIP 269  
 Db 122 VAEIASIP 129

RESULT 2  
 MSH6\_ARATH  
 ID MSH6\_ARATH STANDARD; PRT; 1324 AA.  
 AC O04716;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA mismatch repair protein MSH6-1 (AtMSH6-1).  
 GN MSH6-1 OR AGAA.3 OR AT4G02070 OR T10M13.8.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Landsberg erecta;  
 RA Till S., Granat S., Parnell L., Kaplan N., Hoffman J., Lodhi M.,  
 RA Johnson A.F., Dedhia N., Martienssen R., McCombie W.R.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=99444907; PubMed=10517319;  
 RA Ade J., Beizile F., Philippe H., Doutriaux M.P.;  
 RT "Four mismatch repair paralogues coexist in Arabidopsis thaliana:  
 AtMSH2, AtMSH3, AtMSH6-1 and AtMSH6-2."  
 RL Mol. Gen. Genet. 262:239-249(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=20083488; PubMed=10617198;  
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,  
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,  
 RA Harris B., Ansong W., Brandt P., Grivell L.A., Rieger M.,  
 RA Weichselgartner M., de Simone V., Obermaier B., Macho R., Mueller M.,  
 RA Kreis M., Delzeny M., Puigdomenech P., Watson M., Wachter T.,  
 RA Reichert B., Portetle D., Perez-Alonso M., Boutry M., Schmidheini T.,  
 RA Vos P., Hohlseil J., Zimmermann W., Wedler H., Ridley P.,  
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,  
 RA Van der Schueren J., Ghymonprez B., Chuang Y.-J., Vandenbussche F.,  
 RA Braeken M., Weljens I., Voet M., Bastiaens I., Aert R., Defoor E.,  
 RA Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,  
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirks W.,  
 RA Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,  
 RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,  
 RA De Keyser A., Buyssehaert C., Gielen J., Villarroel R., De Clercq R.,  
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,  
 RA Pettett A., Rajandream M.A., Lyne M., Benes V., Reckmann S.,  
 RA Borkova D., Bloeker H., Scharfe M., Grimm M., Lochner T.-H.,  
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,



```
RA Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
RA Schnabl S., Hiller F., Schmidt W., Lecharny A., Aubourg S.,
RA Chedof F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Mix P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Papan K., Hillier L.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RL thaliana";
RL Nature 402:769-777(1999).
CC -1- FUNCTION: INVOLVED IN POST-REPLICATIVE DNA-MISMATCH REPAIR (BY
CC SIMILARITY).
CC -1- SUBUNIT: HETERODIMER OF MSH2 AND MSH6 (GTBP) (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF001535; AAB57798.1; -
CC EMBL; AJ245967; CAB53337.1; -
CC EMBL; AF001308; AAC78699.1; -
CC EMBL; AL161493; CAB80700.1; -
CC InterPro; IPR000432; Muts_C.
CC InterPro; IPR002863; Muts_N.
CC Pfam; PF00488; Muts_C; 1.
CC Pfam; PF01624; Muts_N; 2.
CC ProDom; PD001263; Muts_C; 1.
CC SMART; SM00534; Mutsac; 1.
CC SMART; SM00533; Mutsd; 1.
CC SMART; SM00333; Tudor; 1.
CC PROSITE; PS00486; DNA_MISMATCH_REPAIR_2; 1.
CC DNA repair; ATP-binding; DNA-binding.
CC NP_BIND; 1083; 1090; ATP (POTENTIAL).
CC CONFLICT; 316; 318; QDR -> GKPKSLGL (IN REF. 1).
CC CONFLICT; 383; 383; Q -> QVRRAHGL (IN REF. 1).
CC CONFLICT; 852; 852; P -> PGINLLHILILANCTASHIISLP (IN
CC REF. 1).
CC
CC SEQUENCE 1324 AA; 146797 MW; 787A340272CF979C CRC64;
Query Match 2.7%; Score 8; DB 1; Length 1324;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 19 SLPNISS 26
DB 855 SLPNISS 862
RESULT 3
K2M1 SHEEP
ID K2M1 SHEEP PRT; 109 AA.
AC P02539;
```

```
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 26-OCT-2001 (Rel. 40, Last annotation update)
DE Keratin, type II microfibrillar (Low-sulfur keratin) (Fragment).
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC NCBI_TaxID=9940;
RN [1]
RP SEQUENCE.
RA MEDLINE=79020757; PubMed=581264;
RA Crewther W.G., Inglis A.S., McKern N.M.;
RT "Amino acid sequences of alpha-helical segments from S-
RT carboxymethylkeratine-A. Complete sequence of a type-II segment.";
RL Biochem. J. 173:365-371(1978).
CC -1- FUNCTION: WOOL MICROFIBRILLAR KERATIN.
CC -1- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC MICROFIBRILLAR KERATIN, I AND II (40-55 AND 56-70 KILODALTONS,
CC RESPECTIVELY). AT LEAST ONE TYPE OF CYTOSKELETAL KERATIN IS
CC PRESENT IN ALL VERTEBRATE EPITHELIAL CELLS.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
DR PIR; A02954; KRSHL2.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament; 1.
DR PROSITE; PS00226; IF; PARTIAL.
KW Intermediate filament; Coiled coil; Keratin.
FT NON TER 1 1
FT DOMAIN <1 >109 ROD.
FT DOMAIN 1 10 LINKER 1.
FT DOMAIN 11 >109 COIL 1B.
FT VARIANT 51 51 R -> K.
FT VARIANT 52 52 R -> K.
FT NON TER 109 109
SQ SEQUENCE 109 AA; 12683 MW; 679DDDA5A621CF8 CRC64;
Query Match 2.4%; Score 7; DB 1; Length 109;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 69 ALKKQVD 75
DB 69 ALKKQVD 75
RESULT 4
VATL DICDI
ID VATL DICDI STANDARD; PRT; 196 AA.
AC P54642;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vacuolar ATP synthase proteolipid subunit (EC 3.6.3.14).
GN VATP.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OC NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX2;
RX MEDLINE=96435786; PubMed=8838672;
RA Xie Y., Coukell M.B., Gombos Z.;
RT "Antisense RNA inhibition of the putative vacuolar H(+)-ATPase
RT proteolipid of Dictyostelium reduces intracellular Ca2+ transport and
RT cell viability";
RL J. Cell Sci. 109:489-497(1996).
CC -1- FUNCTION: PROTON-CONDUCTING PORE FORMING SUBUNIT OF THE MEMBRANE
CC INTEGRAL V0 COMPLEX OF VACUOLAR ATPASE. V-ATPASE IS RESPONSIBLE
CC FOR ACIDIFYING A VARIETY OF INTRACELLULAR COMPARTMENTS IN
CC EUKARYOTIC CELLS.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -1- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A
```

CC PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B,  
 CC C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE  
 CC COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN; WHICH IS PRESENT  
 CC AS A HEXAMER THAT FORMS THE PROTON-CONDUCTING PORE).  
 CC  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Vacuolar.  
 CC  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSION IS MAINTAINED AT A RELATIVELY  
 CC CONSTANT LEVEL DURING GROWTH AND DEVELOPMENT.  
 CC  
 CC -1- MISCELLANEOUS: THIS SUBUNIT BINDS DICICLOHEXYLCARBODIIMIDE (DCDD)  
 CC WHICH INHIBITS THE ATPASE (BY SIMILARITY).  
 CC  
 CC -1- SIMILARITY: BELONGS TO THE V-ATPASE PROTEOLIPID SUBUNIT FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X90516; CAA62102.1; -;  
 CC DictyDB; DD05085; vatP.  
 CC InterPro; IPR002379; ATPase Csub.  
 CC Pfam; PF00137; ATP-synt C; 2.  
 CC PRINTS; PR00122; VACATPASE.  
 CC TIGRFAms; TIGR01100; V\_ATP\_synt C; 1.  
 CC KW Hydrolyase; Hydrogen ion transport; ATP synthesis; Transmembrane.  
 CC FT DOMAIN 1 25 LUMENAL (POTENTIAL).  
 CC FT TRANSMEM 26 46 POTENTIAL.  
 CC FT DOMAIN 47 72 CYTOPLASMIC (POTENTIAL).  
 CC FT TRANSMEM 73 93 POTENTIAL.  
 CC FT DOMAIN 94 111 LUMENAL (POTENTIAL).  
 CC FT TRANSMEM 112 132 POTENTIAL.  
 CC FT DOMAIN 133 150 CYTOPLASMIC (POTENTIAL).  
 CC FT TRANSMEM 151 171 POTENTIAL.  
 CC FT DOMAIN 172 196 LUMENAL (POTENTIAL).  
 CC FT BINDING 158 158 DICICLOHEXYLCARBODIIMIDE (POTENTIAL).  
 CC SQ SEQUENCE 196 AA; 20149 MW; D0FE9438C2F788AB CRC64;  
 CC  
 CC Query Match 2.4%; Score 7; DB 1; Length 196;  
 CC Best Local Similarity 100.0%; Pred. No. 17;  
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 83 ASVGISN 89  
 CC |||||  
 CC DB 54 ASVGISN 60  
 CC  
 CC RESULT 5  
 CC GLXR\_ECOLI STANDARD; PRT; 292 AA.  
 CC ID GLXR\_ECOLI  
 CC AC P77161;  
 CC DT 01-NOV-1997 (Rel. 35, Created)  
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE 2-hydroxy-3-oxopropionate reductase (EC 1.1.1.60) (Tartronate  
 CC semialdehyde reductase) (TSAR).  
 CC GN GLXR OR GLXB1 OR B0509.  
 CC OS Escherichia coli.  
 CC OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC EC Escherichia.  
 CC OX NCBI\_TaxID=562;  
 CC [1]  
 CC RN SEQUENCE FROM N.A.  
 CC RC STRAIN=K12 / MG1655;  
 CC RX MEDLINE=97426617; PubMed=9278503;  
 CC RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 CC Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 CC Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 CC Mau B., Shao Y.;  
 CC RT "The complete genome sequence of Escherichia coli K-12.";  
 CC RL Science 277:1453-1474(1997).  
 CC RN [2]

RP SEQUENCE FROM N.A.  
 RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,  
 RA Duncan M., Federspiel N., Hyman R., Kaiman S., Komp C., Kurdi O.,  
 RA Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 CC [3]  
 CC SEQUENCE FROM N.A., AND FUNCTION.  
 CC STRAIN=K12 / ECL1;  
 CC MEDLINE=20069628; PubMed=10601204;  
 CC RX Cusa E., Obradors N., Baldoma L., Badia J., Aguilar J.;  
 CC RA "Genetic analysis of a chromosomal region containing genes required  
 CC for assimilation of allantoin nitrogen and linked glyoxylate  
 CC metabolism in Escherichia coli.";  
 CC RT J. Bacteriol. 181:7479-7484(1999).  
 CC RL J. Bacteriol. 181:7479-7484(1999).  
 CC CC -1- CATALYTIC ACTIVITY: (R)-glycerate + NAD(P)(+) = 2-hydroxy-3-  
 CC oxopropanoate + NAD(P)H.  
 CC CC -1- PATHWAY: GLYOXYLATE CATABOLISM.  
 CC CC -1- INDUCTION: BY GLYOXYLATE.  
 CC CC -1- SIMILARITY: BELONGS TO THE 3-HYDROXYISOBUTYRATE DEHYDROGENASE  
 CC FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AE000157; AAC73611.1; -;  
 CC EMBL; U82664; AAB40262.1; -;  
 CC EMBL; U89279; AAB93851.1; -;  
 CC EcoGene; EGI3265; glxR.  
 CC InterPro; IPR002204; 3hydroxisobut\_dh.  
 CC DR InterPro; IPR001744; 6PGD.  
 CC DR Pfam; PF03446; NAD binding 2; 1.  
 CC DR PROSITE; PS00895; 3-HYDROXYISOBUT DH; 1.  
 CC KW Oxidoreductase; NAD; Complete proteome.  
 CC FT ACT\_SITE 169 169 BY SIMILARITY.  
 CC SQ SEQUENCE 292 AA; 30800 MW; 5D5263231F2910FO CRC64;  
 CC  
 CC Query Match 2.4%; Score 7; DB 1; Length 292;  
 CC Best Local Similarity 100.0%; Pred. No. 24;  
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 284 GGEIGAR 290  
 CC |||||  
 CC DB 121 GGEIGAR 127  
 CC  
 CC RESULT 6  
 CC FPG\_BRUME STANDARD; PRT; 293 AA.  
 CC ID FPG\_BRUME  
 CC AC Q8VED2;  
 CC DT 15-JUN-2002 (Rel. 41, Created)  
 CC DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 CC DE Formamidopyrimidine-DNA glycosylase (EC 3.2.2.23) (Fapy-DNA  
 CC glycosylase).  
 CC GN MUTM OR FPG OR BME11946.  
 CC OS Brucella melitensis.  
 CC OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 CC EC Brucellaceae; Brucella.  
 CC OX NCBI\_TaxID=29459;  
 CC [1]  
 CC RN SEQUENCE FROM N.A.  
 CC RC STRAIN=16M / ATCC 23456 / Biotype 1;  
 CC RX MEDLINE=20020109; PubMed=11756588;  
 CC RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mijer C., Los T.,  
 CC Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,  
 CC Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,  
 CC Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,  
 CC Haselkorn R., Kyrpides N., Overbeek R.;

RT "The genome sequence of the facultative intracellular pathogen  
 RL Brucella melitensis."; U.S.A. 99:443-448 (2002).  
 CC -!- FUNCTION: This enzyme may play a significant role in processes  
 CC leading to recovery from mutagenesis and/or cell death by  
 CC alkylating agents (By similarity).  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of DNA containing ring-opened N(7)-  
 CC methylguanine residues, releasing 2,6-diamino-4-hydroxy-5-(N-  
 CC methyl)formamidopyrimidine.  
 CC -!- COFACTOR: Binds 1 zinc ion (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE FPG FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AE009628; AAL53127.1; -  
 CC InterPro; IPR000214; Fapy\_DNAGlyco\_zn.  
 CC InterPro; IPR000191; Fapy\_DNA\_glyco.  
 CC Pfam; PF01149; Fapy\_DNA\_glyco.1.  
 CC ProDom; PD003680; Fapy\_DNA\_glyco; 1.  
 CC TIGRFAMs; TIGR00577; fpg; 1.  
 CC PROSITE; PS01242; FPG; 1.  
 CC DNA repair; Hydrolase; Glycosidase; Zinc; Zinc-finger;  
 CC Complete proteome.  
 CC ZN\_FING 266  
 CC SEQUENCE 293 AA; 32005 MW; 2AF0278696CFB7B CRC64;  
 CC  
 CC Query Match 2.4%; Score 7; DB 1; Length 293;  
 CC Best Local Similarity 100.0%; Pred. No. 24;  
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 177 DGLSIIS 183  
 CC 66 DGLSIIS 72  
 CC  
 CC RESULT 7  
 CC HEMZ\_CAMJE STANDARD; PRT; 303 AA.  
 CC ID\_HENZ\_CAMJE  
 CC AC Q9PI08;  
 CC DT 16-OCT-2001 (Rel. 40, Created)  
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE Ferrochelatase (EC 4.99.1.1) (Prothème ferro-lyase) (Heme  
 CC synthetase).  
 CC GN HEMH OR C00503.  
 CC OS Campylobacter jejuni.  
 CC OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
 CC OC Campylobacter.  
 CC OX NCBI\_TaxID=197;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=NCTC 11168;  
 CC MEDLINE=20150912; PubMed=10688204;  
 CC RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
 CC RA Braham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,  
 CC RA Jørgensen K., Kariyasev A.V., Moule S., Pallen M.J., Penn C.W.,  
 CC RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
 CC RA Whitehead S., Barrell B.G.;  
 CC RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
 CC reveals hypervariable sequences".  
 CC RL Nature 403:665-668 (2000).  
 CC -!- FUNCTION: CATALYZES THE FERROUS INSERTION INTO PROTOPORPHYRIN IX.  
 CC -!- CATALYTIC ACTIVITY: Protoporphyrin + Fe(2+) = protoheme + 2 H(+).  
 CC -!- PATHWAY: Protoheme biosynthesis; last step.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE FERROCHELATASE FAMILY.  
 CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AL139075; CAB75140.1; -  
 CC InterPro; IPR001015; Ferrochelatase.  
 CC Pfam; PF00762; Ferrochelatase; 1.  
 CC ProDom; PD002792; Ferrochelatase; 1.  
 CC TIGRFAMs; TIGR00109; hemH; 1.  
 CC PROSITE; PS00534; FERROCHELATASE; 1.  
 CC Porphyrin biosynthesis; Heme biosynthesis; Lyase; Iron;  
 CC Complete proteome.  
 CC FT METAL 185 185 IRON (BY SIMILARITY).  
 CC FT METAL 262 262 IRON (BY SIMILARITY).  
 CC SQ SEQUENCE 303 AA; 35202 MW; 3F7C864A5B833C50 CRC64;  
 CC  
 CC Query Match 2.4%; Score 7; DB 1; Length 303;  
 CC Best Local Similarity 100.0%; Pred. No. 25;  
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 223 AYQSKLG 229  
 CC 223 AYQSKLG 229  
 CC  
 CC RESULT 8  
 CC IDH2\_KLUJA STANDARD; PRT; 368 AA.  
 CC ID\_IDH2\_KLUJA  
 CC AC 094230;  
 CC DT 16-OCT-2001 (Rel. 40, Created)  
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 CC DE Isocitrate dehydrogenase [NAD] subunit 2, mitochondrial precursor  
 CC (EC 1.1.1.41) (isocitric dehydrogenase) (NAD+-specific IDH).  
 CC GN IDH2.  
 CC OS Kluyveromyces fragilis (Yeast).  
 CC OC Eukaryota; Fungi; Ascomycota; Saccharomycetaceae; Kluyveromyces;  
 CC OC Saccharomycetaceae; Saccharomycetaceae; Kluyveromyces.  
 CC OX NCBI\_TaxID=28985;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=JBD100;  
 CC MEDLINE=20426969; PubMed=10975257;  
 CC RA Eizinga S.D.J., van Oosterum K., Maat C., Grivell L.A.,  
 CC RA van der Spek H.;  
 CC RT "Isolation and RNA-binding analysis of NAD+-isocitrate  
 CC dehydrogenases from Kluyveromyces fragilis and Schizosaccharomyces  
 CC pombe".  
 CC RL Curr. Genet. 38:87-94 (2000).  
 CC -!- FUNCTION: PERFORMS AN ESSENTIAL ROLE IN THE OXIDATIVE FUNCTION OF  
 CC THE CITRIC ACID CYCLE (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: Isocitrate + NAD(+) = 2-oxoglutarate + CO(2) +  
 CC NADH.  
 CC -!- SUBUNIT: OCTAMER OF TWO NONIDENTICAL SUBUNITS IDH1 AND IDH2 (BY  
 CC SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Mitochondrial (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE ISOCITRATE AND ISOPROPYLALATE  
 CC DEHYDROGENASES FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF045154; AAC69609.1; -  
 CC HSP; P00351; IXAA.

DR InterPro; IPR001804; Isoadh.  
DR InterPro; IPR004434; Mito\_nad\_idh.  
DR Pfam; PF00180; isoadh; 1.  
DR TIGRFAMs; TIGR001175; mito\_nad\_idh; 1.  
DR PROSITE; PS00470; IDH IMDH; 1.  
KW Oxidoreductase; NAD; Tricarboxylic acid cycle; Transit peptide;  
FW Mitochondrion.  
FT TRANSIT 1 14 MITOCHONDRION (BY SIMILARITY).  
FT CHAIN 15 368 ISOCITRATE DEHYDROGENASE [NAD] SUBUNIT 2.  
FT ACT SITE 112 112 BINDING TO ISOCITRATE (BY SIMILARITY).  
SQ SEQUENCE 368 AA; 39579 MW; E6BFAE7F676F5FB6 CRC64;  
Query Match 2.4%; Score 7; DB 1; Length 368;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 65 KNLIALK 71  
DB 93 KNLIALK 99  
RESULT 9  
K2M2\_SHEEP STANDARD; PRT; 491 AA.  
AC PF1241;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Keratin, type II microfibrillar, component 7C.  
OS Ovis aries (Sheep).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Caprinae; Ovis.  
OX NCBI\_TaxID=9940;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=90026244; PubMed=2803231;  
RA Sparrow L.G., Robinson C.P., McMahon D.T.W., Rubira M.R.;  
RT "The amino acid sequence of component 7c, a type II intermediate-  
RT filament protein from wool.";  
RT Biochem. J. 261:1015-1022(1989).  
CC -|- FUNCTION: WOOL MICROFIBRILLAR KERATIN.  
CC -|- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND  
CC MICROFIBRILLAR KERATIN, I AND II (40-55 AND 56-70 KILODALTONS,  
CC RESPECTIVELY). AT LEAST ONE TYPE OF CYTOSKELETAL KERATIN IS  
CC PRESENT IN ALL VERTEBRATE EPITHELIAL CELLS.  
CC -|- MISCELLANEOUS: THE LOW-SULFUR PROTEINS, DERIVED FROM THE  
CC MICROFIBRILLAR FRACTION OF WOOL EXTRACTS, ARE COMPOSED OF TWO  
CC FAMILIES, EACH CONSISTING OF 4 HOMOLOGOUS SUBUNITS: THE TYPE I  
CC COMPONENTS (8C-1, 8C-2, 8A & 8B) AND THE TYPE II COMPONENTS (5,  
CC 7A, 7B AND 7C).  
CC -|- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
DR PIR; S05408; S05408.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR003054; Keratin\_II.  
DR Pfam; PF00038; Keratin; 1.  
DR PRINTS; PR01276; TYPE2KERATIN.  
DR PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Coiled coil; Keratin.  
FT MOD RES 1 1 BLOCKED.  
FT FT DOMAIN 1 109 HEAD.  
FT FT DOMAIN 110 416 ROD.  
FT FT DOMAIN 417 491 TAIL.  
FT FT DOMAIN 110 144 COIL 1A.  
FT FT DOMAIN 145 154 LINKER 1.  
FT FT DOMAIN 155 255 COIL 1B.  
FT FT DOMAIN 256 272 LINKER 12.  
FT FT DOMAIN 273 416 COIL 2.  
FT FT VARIANT 74 74 C -> G OR S.  
FT FT VARIANT 80 80 C -> S.  
FT FT VARIANT 144 144 F -> Y.  
FT FT VARIANT 232 232 S -> V.  
FT FT VARIANT 276 276 C -> D OR N.  
FT FT VARIANT 276 276



OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87254239; PubMed=2439698;  
RA Klinge E.M., Sylvestre Y.R., Freedberg I.M., Blumenberg M.;  
RT "Evolution of keratin genes: different protein domains evolve by  
RT different pathways";  
RL J. Mol. Evol. 24:319-329(1987).  
[2]  
RN VARIANT MCD LYS-509.  
RP MEDLINE=97315826; PubMed=9171831;  
RX Irvine A.D., Corden L.D., Swenson O., Swenson B., Moore J.E.,  
RA Frazer D.G., Smith F.J.D., Knowlton R.G., Christophers E.,  
RA Rochels R., Uitto J., McLean W.H.I.;  
RT "Mutations in cornea-specific keratin K3 or K12 genes cause  
RT Meesmann's corneal dystrophy";  
RL Nat. Genet. 16:184-187(1997).  
CC -!- SURUNIT: HETEROETIMER OF TWO TYPE I AND TWO TYPE II KERATINS.  
CC -!- KERATIN 3 ASSOCIATES WITH KERATIN 12.  
CC -!- TISSUE SPECIFICITY: CORNEA-SPECIFIC.  
CC -!- DISEASE: DEFECTS IN KRT3 ARE A CAUSE OF MEESMANN CORNEAL DYSTROPHY  
CC (MCD), AN AUTOSOMAL DOMINANT DISEASE THAT CAUSES FRAGILITY OF THE  
CC ANTERIOR CORNEAL EPITHELIUM. PATIENTS ARE USUALLY ASYMPTOMATIC  
CC UNTIL ADULTHOOD WHEN RUPTURE OF THE CORNEAL MICROCYSTS MAY CAUSE  
CC EROSIONS, PRODUCING CLINICAL SYMPTOMS SUCH AS PHOTOPHOBIA, CONTACT  
CC LENS INTOLERANCE AND INTERMITTENT DIMINUTION OF VISUAL ACUITY.  
CC RARELY, SUBEPITHELIAL SCARRING CAUSES IRREGULAR CORNEAL  
CC ASTIGMATISM AND PERMANENT VISUAL IMPAIRMENT. HISTOLOGICAL  
CC EXAMINATION SHOWS A DISORGANIZED AND THICKENED EPITHELIUM WITH  
CC WIDESPREAD CYTOPLASMIC VACUOLATION AND NUMEROUS SMALL, ROUND,  
CC DERIS-LADEN INTRAEPITHELIAL CYSTS.  
CC -!- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND  
CC MICROFIBRILLAR KERATIN: I (ACIDIC; 40-55 kDa) [K9 TO K20] AND II  
CC (NEUTRAL TO BASIC; 56-70 kDa) [K1 TO K8].  
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X05418; CAA28991.1; -  
CC EMBL; X05419; CAA28992.1; ALT SEQ.  
CC EMBL; X05420; CAA28993.1; ALT SEQ.  
CC EMBL; X05420; CAA28994.1; ALT SEQ.  
CC EMBL; X05420; CAA28995.1; ALT SEQ.  
CC EMBL; X05421; CAA28996.1; ALT\_SEQ.  
CC PIR; A29666; A29666.  
CC Genew; HGNC:6440; KRT3.  
CC MIM; 148043; -  
CC MIM; 122100; -  
CC InterPro; IPR001664; IF.  
CC InterPro; IPR003054; Keratin\_II.  
CC Pfam; PF00038; filament; 1.  
CC PRINTS; PR01276; TYPE2KERATIN.  
CC PROSITE; PS00226; IF; 1.  
CC Intermediate filament; Coiled coil; Keratin; Phosphorylation;  
KW Disease mutation.  
FT DOMAIN 1 197  
FT DOMAIN 198 509  
FT DOMAIN 510 629  
FT DOMAIN 510 233  
FT COIL 1A.  
FT LINKER 1. 234 254  
FT COIL 1B.  
FT DOMAIN 255 346  
FT DOMAIN 347 370  
FT LINKER 12.  
FT COIL 2.  
FT DOMAIN 371 509  
FT MOD\_RES 56 56  
FT VARIANT 509 509  
E -> K (IN MCD).  
/FTID=VAR\_003868.  
  
SQ SEQUENCE 629 AA; 64511 MW; 2E748619A828BCD9 CRC64;  
Query Match 2.4%; Score 7; DB 1; Length 629;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 70 LKQDVS 76  
DB 307 LKQDVS 313  
[1]  
[2]  
RESULT 14  
GUND\_HAEIN  
ID GUND\_HAEIN STANDARD; PRT; 863 AA.  
AC P43919;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE [Protein-Pil] uridylyltransferase (EC 2.7.7.59) (PII uridylyl-  
DE transferase) (Uridylyl removing enzyme) (UTase).  
GN GLND OR H1719.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RD / KW20 / ATCC 51907;  
RX MEDLINE=9550630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagan N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
RT Rd.";  
RL Science 269:496-512(1995).  
CC -!- FUNCTION: MODIFIES, BY URIDYLYLATION OR DEURIDYLYLATION THE PII  
CC (GLNB) REGULATORY PROTEIN (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: UTP + [protein-PII] = diphosphate + uridylyl-  
CC [protein-PII].  
CC -!- SIMILARITY: BELONGS TO THE GUND FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U32844; AAC23362.1; -  
CC TIGR; H1719; -  
CC InterPro; IPR002912; ACT.  
CC InterPro; IPR002819; HD.  
CC InterPro; IPR003607; ME Pplase HDC.  
CC InterPro; IPR002934; NTP\_transf.  
CC Pfam; PF01842; ACT; 2.  
CC Pfam; PF01909; NTP\_transf\_2; 1.  
CC Pfam; PF01966; HD; 1.  
CC SMART; SM00471; HDC; 1.  
CC Transferrase; Nucleotidyltransferase; Complete proteome.  
SQ SEQUENCE 863 AA; 100173 MW; B6F679A70A01A651 CRC64;  
Query Match 2.4%; Score 7; DB 1; Length 863;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 239 SLFASLY 245

```
Db          590 SLFASLY 596
|||||
RESULT 15
FOX2 YEAST
ID_FOX2 YEAST STANDARD; PRT; 900 AA.
AC Q02207;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Peroxisomal hydratase-dehydrogenase-epimerase (HDE) (Multifunctional
DE beta-oxidation protein) (MFP) [Includes: 2-enoyl-CoA hydratase
DE (EC 4.2.1.-); D-3-hydroxyacyl CoA dehydrogenase (EC 1.1.1.-)].
GN FOX2 OR YKR009C OR YK108.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RA Duesterhoeft A., Philippsen P.;
RX "DNA sequencing and analysis of a 24.7 kb segment encompassing
RT centromere CENII of Saccharomyces cerevisiae reveals nine previously
RT unknown open reading frames.";
RL Yeast 8:749-759(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=92202210; PubMed=1551874;
RA Hiltunen J.K., Wenzel B., Beyer A., Erdmann R., Fossa A., Kunau W.H.;
RX "Peroxisomal multifunctional beta-oxidation protein of Saccharomyces
RT cerevisiae. Molecular analysis of the fox2 gene and gene product.";
RL J. Biol. Chem. 267:6646-6653(1992).
CC -1- FUNCTION: SECOND TRIFUNCTIONAL ENZYME ACTING ON THE BETA-OXIDATION
CC PATHWAY FOR FATTY ACIDS, POSSESSING HYDRATASE-DEHYDROGENASE-
CC EPIMERASE ACTIVITIES. CONVERTS TRANS-2-ENOYL-CoA VIA D-3-
CC HYDROXYACYL-CoA TO 3-KETOACYL-CoA.
CC -1- PATHWAY: BETA-OXIDATION PATHWAY.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Peroxisomal.
CC -1- DOMAIN: CONTAINS TWO SDR DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
EMBL; M86456; AAA34779.1; -
EMBL; X65124; CAA46243.1; -
EMBL; Z28234; CAA82079.1; -
PIR; S25322; S25322.
DR HSSP; O70351; 1B6W.
DR SGD; S0001717; FOX2.
DR InterPro; IPR002198; ADH short.
DR InterPro; IPR002539; MaoC dehydratas.
DR Pfam; PF00106; adh_short; 2.
DR Pfam; PF01575; MaoC dehydratas; 1.
DR PRINTS; PR00080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; 2.
DR PROSITE; PS00342; MICROBODIES_CTR; 1.
KW Fatty acid metabolism; Multifunctional enzyme; Oxidoreductase; NAD;
KW Lyase; Isomerase; Peroxisome; Repeat.
FT DOMAIN 6 230 SHORT-CHAIN DEHYDROGENASE LIKE 1.
FT DOMAIN 319 535 SHORT-CHAIN DEHYDROGENASE LIKE 2.
FT SITE 898 900 MICROBODY TARGETING SIGNAL (POTENTIAL).
FT NP_BIND 13 37 NAD (BY SIMILARITY).
-----
FT ACT_SITE 165 165 BY SIMILARITY.
FT NP_BIND 326 350 NAD (BY SIMILARITY).
FT ACT_SITE 469 469 BY SIMILARITY.
SQ SEQUENCE 900 AA; 98703 MW; 66FFD0D49C673788 CRC64;
Query Match 2.4%; Score 7; DB 1; Length 900;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 164 KVSNSIF 170
Db 232 KVSNSIF 238
RESULT 16
YEGO ECOLI
ID_YEGO ECOLI STANDARD; PRT; 1025 AA.
AC P76399; O08006;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yego.
GN YEGO OR B2076.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RA Blattner F.R., Plunkett G. III, Mayhew G.F., Perna N.T.,
RX Glatner F.D.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12; PubMed=9097040;
RX MEDLINE=97251358;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saito N., Sempel G., Seki Y., Sivasubram S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
CC -1- FUNCTION: COULD BE A DRUG EFFLUX PUMP (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (potential).
CC -1- SIMILARITY: BELONGS TO THE ACRB/ACRD/ACRF FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
EMBL; AE000297; AAC75137.1; -
EMBL; D90845; BAA15930.1; -
EMBL; D90846; BAA15932.1; -
EGene; EG14058; Yego.
DR InterPro; IPR001036; Acrflvin_res.
DR Pfam; PF00873; ACR_tran; 1.
DR PRINTS; PR00702; ACRIFLAVINRP.
DR Hypothetical protein; Transmembrane; Inner membrane; Transport;
KW Complete proteome.
FT DOMAIN 1 11 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 12 30 1 (BY SIMILARITY).
FT DOMAIN 31 333 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 334 353 2 (BY SIMILARITY).
FT DOMAIN 354 359 CYTOPLASMIC (BY SIMILARITY).
```

```
FT TRANSMEM 360 379 3 (BY SIMILARITY).
FT DOMAIN 380 385 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 386 407 4 (BY SIMILARITY).
FT DOMAIN 408 434 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 435 453 5 (BY SIMILARITY).
FT DOMAIN 454 466 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 467 489 6 (BY SIMILARITY).
FT DOMAIN 490 526 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 527 545 7 (BY SIMILARITY).
FT DOMAIN 546 851 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 852 871 8 (BY SIMILARITY).
FT DOMAIN 872 877 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 878 897 9 (BY SIMILARITY).
FT DOMAIN 898 903 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 904 925 10 (BY SIMILARITY).
FT DOMAIN 926 953 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 954 972 11 (BY SIMILARITY).
FT DOMAIN 973 985 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 986 1008 12 (BY SIMILARITY).
FT DOMAIN 1009 1025 CYTOPLASMIC (BY SIMILARITY).
SQ SEQUENCE 1025 AA; 111010 MW; EF00BB4E7B301008 CRC64;

Query Match 2.4%; Score 7; DB 1; Length 1025;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 FSLNNLS 198
DB 382 FSLNNLS 388

RESULT 17
SBCC_LACIA STANDARD; PRT; 1046 AA.
AC Q9CFZ0;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Surfactin sbccD subunit C.
GN SBCC OR L14321.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,
RA Weisenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis Il1403."
RL Genome Res. 11:731-753(2001).
CC -!- FUNCTION: SbcD cleaves DNA hairpin structures. These structures
CC can inhibit DNA replication and are intermediates in certain DNA
CC recombination reactions. The complex acts as a 3'->5' double
CC strand exonuclease that can open hairpins. It also has a 5'
CC single-strand endonuclease activity (By similarity).
CC -!- SUBUNIT: Heterodimer of sbcC and sbcD (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. SBCC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE006364; AK05419.1; -
CC InterPro; IPR003439; ABC transport.
KW Hydrolyase; Nuclease; Exonuclease; Endonuclease; DNA replication;
KW DNA recombination; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 34 41
ATP (POTENTIAL).
```

```
FT DOMAIN 223 239 COILED COIL (POTENTIAL).
FT DOMAIN 268 432 COILED COIL (POTENTIAL).
FT DOMAIN 468 505 COILED COIL (POTENTIAL).
FT DOMAIN 543 867 COILED COIL (POTENTIAL).
SQ SEQUENCE 1046 AA; 120199 MW; 8F70D00AC28F8691 CRC64;

Query Match 2.4%; Score 7; DB 1; Length 1046;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 NQFNLN 258
DB 280 NQFNLN 286

RESULT 18
SRF2_BACSU STANDARD; PRT; 3587 AA.
AC Q04747;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Surfactin synthetase subunit 2.
GN SRFAB OR SRF2A OR COML.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE OF 1-3077 FROM N.A.
RC STRAIN=168;
RX MEDLINE=93181186; PubMed=8441623;
RA Puma S., Fujishima Y., Corbell N., D'Souza C., Nakano M.M.,
RA Zuber P., Yamane K.;
RT "Nucleotide sequence of 5' portion of srfA that contains the region
RT required for competence establishment in Bacillus subtilis."
RL Nucleic Acids Res. 21:93-97(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=93360813; PubMed=8355609;
RA Cosmina P., Rodriguez F., de Ferra F., Grandi G., Perego M.,
RA Venema G., van Sinderen D.;
RT "Sequence and analysis of the genetic locus responsible for surfactin
RT synthesis in Bacillus subtilis."
RL Mol. Microbiol. 8:821-831(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124189; PubMed=8969502;
RA Yanane K., Kumano M., Kurita K.;
RT "The 25 degrees-36 degrees region of the Bacillus subtilis
RT chromosome: determination of the sequence of a 146 kb segment and
RT identification of 113 genes."
RL Microbiology 142:3047-3056(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haesch J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
```





CC OVPOSITION.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: MAIN CELLS OF THE ACCESSORY GLANDS OF MALES  
 CC (PARAGONIAL GLAND).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X99412; CAA67789.1; -;  
 DR EMBL; X99417; CAA67794.1; -;  
 DR FlyBase; FBgn0021369; DmauAcp70A.  
 DR FlyBase; FBgn0021129; DaimAcp70A.  
 KW Behavior; Hydroxylation; Signal.  
 FT SIGNAL 1 19 BY SIMILARITY.  
 FT CHAIN 20 55 ACCESSORY GLAND-SPECIFIC PEPTIDE 70A.  
 FT MOD RES 28 28 HYDROXYLATION (BY SIMILARITY).  
 FT MOD RES 32 32 HYDROXYLATION (BY SIMILARITY).  
 FT MOD RES 34 34 HYDROXYLATION (BY SIMILARITY).  
 FT MOD RES 38 38 HYDROXYLATION (BY SIMILARITY).  
 SQ SEQUENCE 55 AA; 6424 MW; E87397C3F196123D CRC64;  
 Query Match 2.0%; Score 6; DB 1; Length 55;  
 Best Local Similarity 100.0%; Pred. No. 61;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 13 VLVCLL 18  
 Db 9 VLVCLL 14  
 |||||  
 RESULT 20  
 A70A DROSE  
 ID A70A DROSE STANDARD; PRT; 55 AA.  
 AC O18417;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Accessory gland-specific peptide 70A precursor (Paragonial peptide B).  
 GN PABP OR ACP70A.  
 OS Drosophila sechellia (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7238;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97432799; PubMed=9286679;  
 RA Citera S., Aguade M.N.;  
 RT "Evolutionary history of the sex-peptide (Acp70A) gene region in  
 RT Drosophila melanogaster";  
 RL Genetics 147:189-197(1997).  
 CC -!- FUNCTION: REPRESENTS FEMALE SEXUAL RECEPTIVITY AND STIMULATES  
 CC OVPOSITION.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: MAIN CELLS OF THE ACCESSORY GLANDS OF MALES  
 CC (PARAGONIAL GLAND).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X99414; CAA67791.1; -;  
 DR FlyBase; FBgn0012779; DsecAcp70A.  
 KW Behavior; Hydroxylation; Signal.

FT SIGNAL 1 19 BY SIMILARITY.  
 FT CHAIN 20 55 ACCESSORY GLAND-SPECIFIC PEPTIDE 70A.  
 FT MOD RES 28 28 HYDROXYLATION (BY SIMILARITY).  
 FT MOD RES 32 32 HYDROXYLATION (BY SIMILARITY).  
 FT MOD RES 34 34 HYDROXYLATION (BY SIMILARITY).  
 FT MOD RES 38 38 HYDROXYLATION (BY SIMILARITY).  
 SQ SEQUENCE 55 AA; 6438 MW; E87AD0919657A83D CRC64;  
 Query Match 2.0%; Score 6; DB 1; Length 55;  
 Best Local Similarity 100.0%; Pred. No. 61;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 13 VLVCLL 18  
 Db 9 VLVCLL 14  
 |||||  
 RESULT 21  
 PRGI SALTY  
 ID PRGI SALTY STANDARD; PRT; 80 AA.  
 AC P41784;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE PRGI protein.  
 DE PRGI OR STM2873.  
 GN Salmonella typhimurium.  
 OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 14028S;  
 RX MEDLINE=96020671; PubMed=7476203;  
 RA Pegues D.A., Hantman M.J., Behlau I., Miller S.I.;  
 RT "PhoP/PhoQ transcriptional repression of Salmonella typhimurium  
 RT invasion genes: evidence for a role in protein secretion";  
 RL Mol. Microbiol. 17:169-181(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 RT LT2";  
 RL Nature 413:852-856(2001).  
 CC -!- FUNCTION: REQUIRED FOR INVASION OF EPITHELIAL CELLS.  
 CC -!- SIMILARITY: BELONGS TO THE MXI/PRGI/YSCF FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U21676; AAB60189.1; -;  
 DR EMBL; AE008831; AAL21753.1; -;  
 DR StyGene; SGI0545; prgi.  
 KW Virulence; Transport; Protein transport; Complete proteome.  
 SQ SEQUENCE 80 AA; 8857 MW; ED1F8967877585F8 CRC64;  
 Query Match 2.0%; Score 6; DB 1; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 84;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 223 AYQSKL 228  
 |||||

Db 46 AYQSKL 51

RESULT 22

RR15\_PINTH STANDARD; PRT; 88 AA.

AC P41648;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Chloroplast 30S ribosomal protein S15.

GN RPS15.

OS pinus thunbergii (Green pine) (Japanese black pine).

OG Chloroplast.

CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

CC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.

OX NCBI\_TaxID=33350;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95024047; PubMed=7937893;

RA Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Taudzuki T.,

RA Sugitara M.;

RT "Loss of all ndh genes as determined by sequencing the entire

RT chloroplast genome of the black pine *Pinus thunbergii*."

RL Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798 (1994).

CC -1- SIMILARITY: BELONGS TO THE S15P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; D17510; BAA04444.1; -.

DR HSSP; P80378; 1AB3.

DR InterPro; IPR005290; RS15\_bact.

DR InterPro; IPR00589; Ribosomal\_S15.

DR Pfam; PF00312; Ribosomal\_S15; 1.

DR TIGRFAMs; TIGR00952; S15\_bact; 1.

DR PROSITE; PS00362; RIBOSOMAL\_S15; 1.

KW Ribosomal protein; Chloroplast; rRNA-binding.

SQ SEQUENCE 88 AA; 10312 MW; 30FD1FAF4D3A8F3C CRC64;

Query Match 2.0%; Score 6; DB 1; Length 88;

Best Local Similarity 100.0%; Pred. No. 91;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 195 NNLIS 200

Db 3 NNLIS 8

RESULT 23

YL34\_ARCFU STANDARD; PRT; 96 AA.

AC O28146;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein AF2134 precursor.

GN AF2134.

OS Archaeoglobus fulgidus.

CC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;

CC Archaeoglobaceae; Archaeoglobus.

OX NCBI\_TaxID=2234;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;

RX MEDLINE=98049343; PubMed=9389475;

RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,

RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,

Richardson D.L., Kervlavage A.R., Graham D.B., Kyripides N.C.,

Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,

Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,

Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodet A., Zhou L.,

Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,

Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,

Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,

Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,

Venter J.C.;

RT "The complete genome sequence of the hyperthermophilic, sulphate-

RT reducing archaeon *Archaeoglobus fulgidus*."

RL Nature 390:364-370 (1997).

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; AE000957; AAB89126.1; -.

DR TIGR; AF2134; -.

KW Hypothetical protein; Signal; Complete proteome.

FT SIGNAL 1 21 POTENTIAL.

FT CHAIN 22 96 HYPOTHETICAL PROTEIN AF2134.

SQ SEQUENCE 96 AA; 10837 MW; 019A92CAC271EBB5 CRC64;

Query Match 2.0%; Score 6; DB 1; Length 96;

Best Local Similarity 100.0%; Pred. No. 99;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 73 DVDSIE 78

Db 60 DVDSIE 65

RESULT 24

QACG STASP STANDARD; PRT; 107 AA.

AC O87856;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Quaternary ammonium compound-resistance protein qacG (Quaternary

DE ammonium determinant G).

GN QACG.

OS *Staphylococcus* sp. (strain ST94).

OG Plasmid pST94.

OC Bacteria; Firmicutes; Bacillales; Staphylococcus.

OX NCBI\_TaxID=29387;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98295006; PubMed=9631545;

RA Heir E., Sundheim G., Holck A.L.;

RT "The *Staphylococcus* qacG gene product: a new member of the SMR family

RT encoding multidrug resistance."

RL FEMS Microbiol. Lett. 163:49-56 (1998).

CC -1- FUNCTION: MULTIDRUG EXPORTER. IS IMPLICATED FOR THE RESISTANCE TO

CC BACTERIOCIDAL QUATERNARY AMMONIUM COMPOUNDS.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -1- SIMILARITY: BELONGS TO THE SMALL MULTIDRUG RESISTANCE (SMR)

CC PROTEIN FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; Y16944; CAA76542.1; -.

```
DR InterPro: IPR000390; DUF7.
DR Pfam: PF00893; DUF7; 1.
KW Plasmaid; Transmembrane; Transport.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 26 46 POTENTIAL.
FT TRANSMEM 57 77 POTENTIAL.
FT TRANSMEM 84 104 POTENTIAL.
SQ SEQUENCE 107 AA; 11819 MW; DBF5F00B21EAE3BF CRC64;

Query Match      2.0%; Score 6; DB 1; Length 107;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 IISVIV 186
DB 72 IISVIV 77

RESULT 25
Y098_ARCFU
ID Y098_ARCFU STANDARD; PRT; 107 AA.
AC O30138;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF0098.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickley E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirschner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.P., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF001099; AB91132.1; -.
CC TIGR; AF0098; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 9 28 POTENTIAL.
FT TRANSMEM 33 50 POTENTIAL.
FT TRANSMEM 55 72 POTENTIAL.
FT TRANSMEM 77 99 POTENTIAL.
SQ SEQUENCE 107 AA; 11830 MW; 86F967B9232EFOFE CRC64;

Query Match      2.0%; Score 6; DB 1; Length 107;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 SLFASL 244
```

```
DB 43 SLFASL 48
|||||
RESULT 26
KV1H_HUMAN
ID KV1H_HUMAN STANDARD; PRT; 108 AA.
AC P01600;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-I region Hau.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=71032830; PubMed=4097974;
RA Watanabe S., Hilesmann N.;
RT "The primary structure of a monoclonal kappa-type immunoglobulin L-
chain of subgroup I (Bence-Jones Protein Hau): subdivision within
RT subgroups.";
RL Hoppe-Seyler's Z. Physiol. Chem. 351:1291-1295(1970).
CC -!- MISCELLANEOUS: THE C REGION OF THIS CHAIN HAS THE INV (3) MARKER.
CC -!- MISCELLANEOUS: THIS IS A BENCE-JONES PROTEIN.
DR PIR; A01868; K1HUHU.
DR HSSP; P80362; 1WTL.
DR InterPro; IPR003066; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00406; IGv; 1.
KW Immunoglobulin V region; Bence-Jones protein.
FT DOMAIN 1 23 FRAMEWORK-1.
FT DOMAIN 24 34 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 35 49 FRAMEWORK-2.
FT DOMAIN 50 56 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 57 88 FRAMEWORK-3.
FT DOMAIN 89 97 FRAMEWORK-4.
FT DOMAIN 98 107 COMPLEMENTARITY-DETERMINING-3.
FT DISULFID 23 88 BY SIMILARITY.
FT NON TER 108 108
SQ SEQUENCE 108 AA; 11671 MW; 08D3A6160D8D0618 CRC64;

Query Match      2.0%; Score 6; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 GTRVEI 124
|||||
DB 101 GTRVEI 106

RESULT 27
KV3F_HUMAN
ID KV3F_HUMAN STANDARD; PRT; 109 AA.
AC P01624;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ig kappa chain V-III region POM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=76276460; PubMed=60899;
RA Klapper D.G., Capra J.D.;
RT "The amino acid sequence of the variable regions of the light chains
from two idiotypically cross reactive Igm anti-gamma globulins.";
RL Ann. Immunol. (Paris) 127C:261-271(1976).
CC -!- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGM WITH ANTI-GAMMA
```

```

CC      GLOBULIN ACTIVITY.
DR      PIR: A01897; K3HUPM.
DR      HSSP; P80362; 1WTL.
DR      InterPro; IPR003006; Ig MHC.
DR      InterPro; IPR003596; Ig_v.
DR      Pfam; PF00047; Ig; 1.
DR      SMART; SM00406; IGV; 1.
KW      immunoglobulin v region.
FT      DISULFID 23 89
FT      NON TER 109 109
SQ      SEQUENCE 109 AA; 11922 MW; 62821DDC6A8ABA86 CRC64;

Query Match      2.0%; Score 6; DB 1; Length 109;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      119 GTRVEI 124
Db      102 GTRVEI 107
|||||

RESULT 28
V441_BPMD2      STANDARD; PRT; 111 AA.
AC      O64236;
DT      15-DEC-1998 (Rel. 37, Created)
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DT      15-DEC-1998 (Rel. 37, Last annotation update)
DE      Gene 44.1 protein (GP44.1).
GN      44.1.
OS      Mycobacteriophage D29.
OC      Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC      unclassified Siphoviridae.
OX      NCBI_TaxID=28369;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=98300335; PubMed=9636706;
RA      Ford M.E., Sarkis G.J., Belanger A.E., Hendrix R.W., Hatfull G.F.;
RT      "Genome structure of mycobacteriophage D29: implications for phage
RL      evolution.";
RL      J. Mol. Biol. 279:143-164 (1998).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AF022214; AAC18486.1;
CC      SEQUENCE 111 AA; 12154 MW; 1EA280A4E2EE12B5 CRC64;

Query Match      2.0%; Score 6; DB 1; Length 111;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      283 FGGEIG 288
Db      58 FGGEIG 63
|||||

RESULT 29
DAD1_XENLA      STANDARD; PRT; 113 AA.
AC      P45967;
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      15-JUL-1999 (Rel. 38, Last annotation update)
DE      Defender against cell death 1 (DAD-1).
GN      DAD1.
OS      Xenopus laevis (African Clawed frog).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC      Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC      Xenopodinae; Xenopus.
OX      NCBI_TaxID=83355;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=94019310; PubMed=8413235;
RA      Nakashima T., Sekiguchi T., Kuraoka A., Fukushima K., Shibata Y.,
RA      Komiyama S., Nishimoto T.;
RT      "Molecular cloning of a human cDNA encoding a novel protein, DAD1,
RT      whose defect causes apoptotic cell death in hamster BHK21 cells.";
RL      Mol. Cell. Biol. 13:6367-6374 (1993).
CC      -1- FUNCTION: LOSS OF THE DAD1 PROTEIN TRIGGERS APOPTOSIS.
CC      -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC      -1- SIMILARITY: BELONGS TO THE DAD1 / OST2 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; D15059; BAA03652.1;
CC      InterPro; IPR003038; DAD.
CC      Pfam; PF02109; DAD; 1.
CC      Apoptosis; Transmembrane.
FT      TRANSMEM 31 51
FT      TRANSMEM 53 73
SQ      SEQUENCE 113 AA; 12596 MW; B8B4FB18524D59B2 CRC64;

Query Match      2.0%; Score 6; DB 1; Length 113;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      48 SVSVFS 53
Db      2 SVSVFS 7
|||||

RESULT 30
NU3M_PARLI      STANDARD; PRT; 117 AA.
AC      P12774;
DT      01-OCT-1989 (Rel. 12, Created)
DT      01-OCT-1989 (Rel. 12, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      NADH-ubiquinone oxidoreductase chain 3 (EC 1.6.5.3).
GN      ND3.
OS      Paracentrotus lividus (Common sea urchin).
OC      Mitochondrion.
OC      Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC      Echinoidea; Euechinoidea; Echinacea; Echinoidea; Echinidae;
OC      Paracentrotus.
OX      NCBI_TaxID=7656;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=89291831; PubMed=2544576;
RA      Cantatore P., Roberti M., Rainaldi G., Gadaleta M.N., Saccone C.;
RT      "The complete nucleotide sequence, gene organization, and genetic
RT      code of the mitochondrial genome of Paracentrotus lividus.";
RL      J. Biol. Chem. 264:10965-10975 (1989).
CC      -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC      -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 3 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----

```

```
DR EMBL; J04815; AAA68141.2; -
DR PIR; I34284; I34284.
DR InterPro; IPR000440; Oxidored_q4.
DR Pfam; PF00507; oxidored_q4; 1.
DR Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ SEQUENCE 117 AA; 13051 MW; 8FEC8EBD1ED032B7 CRC64;

Query Match 2.0%; Score 6; DB 1; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 ITSVA 276
DB 11 ITSVA 16

RESULT 31
RL20 ANASP STANDARD; PRT; 118 AA.
AC Q8YKL8;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L20.
GN RPLT OR RPL20 OR ALR3428.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA RES. 8:205-213 (2001).
CC -!- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA AND IS
CC NECESSARY TO THE IN VITRO ASSEMBLY PROCESS OF THE 50S RIBOSOMAL
CC SUBUNIT; IT IS NOT INVOLVED IN THE PROTEIN SYNTHESIZING FUNCTIONS
CC OF THAT SUBUNIT (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE L20P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP003593; BAB75127.1; -
CC InterPro; IPR001081; Ribosomal L20.
CC Pfam; PF00453; Ribosomal L20; 1.
CC PRINTS; PR00062; Ribosomal L20.
CC ProDom; PD002389; Ribosomal L20; 1.
CC TIGRFAMs; TIGR01032; rplbact; 1.
CC PROSITE; PS00937; RIBOSOMAL_L20; 1.
CC Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 118 AA; 13505 MW; 09EEB696B1D17B31 CRC64;

Query Match 2.0%; Score 6; DB 1; Length 118;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 VAEAS 267
DB 109 VAEAS 114

RESULT 32
RL20 ANASP STANDARD; PRT; 119 AA.
AC P53726;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 13.7 kDa protein in SWP73-SEC12 intergenic region.
GN YNR025C OR N3235.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Pohl T.M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z71639; CAA96304.1; -
CC SGD; S0005308; YNR025C.
CC Hypothetical protein; Transmembrane.
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 32 52 POTENTIAL.
SQ SEQUENCE 119 AA; 13725 MW; 1D6D0566B0F4D8D5 CRC64;

Query Match 2.0%; Score 6; DB 1; Length 119;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 SLNLS 198
DB 53 SLNLS 58

RESULT 33
VG39 BPMD2 STANDARD; PRT; 128 AA.
AC O64230;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Gene 39 protein (GP39).
GN 39.
OS Mycobacteriophage D29.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC Unclassified Siphoviridae.
OX NCBI_TaxID=28369;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98300335; PubMed=9636706;
RA Ford M.E., Sarkis G.J., Belanger A.E., Hendrix R.W., Hatfull G.F.;
RT "Genome structure of mycobacteriophage D29: implications for phage
RT evolution."
RL J. Mol. Biol. 279:143-164 (1998).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF022214; AAC18480.1; -
CC SEQUENCE 128 AA; 14816 MW; 21D632CB2E6B4A5F CRC64;
```

```
Query Match      2.0%; Score 6; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      213 EFFDVL 218
DB      8 EFFDVL 13

RESULT 34
KV3H HUMAN      STANDARD;      PRT;      129 AA.
AC      P04207.
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DE      Hypothetical 14.2 kDa protein.
DT      15-JUL-1999 (Rel. 38, Last annotation update)
DE      Leuconostoc mesenteroides (subsp. cremoris).
OS      Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OC      NCBI_TaxID=33965;
ON      (1)
RN      SEQUENCE FROM N.A.
RA      Cavin J.F., Dartois V.A., Divies C.;
RL      Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC      -1- SIMILARITY: BELONGS TO THE UPF0076 (UK114) FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; U50749; AAB48552.1; -.
DR      HSSP; P37552; 1QD9.
DR      InterPro; IPR000543; YjgF-like.
DR      Pfam; PF01042; UPF0076; 1.
DR      TIGRFAMs; TIGR00004; YER057c_YjgF_UK114; 1.
DR      PROSITE; PS01094; UPF0076; 1.
KW      Hypothetical protein.
SQ      SEQUENCE 130 AA; 14232 MW; 74531BAD1F2CA525 CRC64;

Query Match      2.0%; Score 6; DB 1; Length 130;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      39 LYISGQ 44
DB      28 LYISGQ 33

RESULT 36
HSLR_ECOLI
ID      HSLR_ECOLI      STANDARD;      PRT;      133 AA.
AC      P45802;
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Heat shock protein 15 (HSP15).
GN      HSLR OR B3400 OR Z4754 OR ECS4242.
OS      Escherichia coli, and
OS      Escherichia coli O157:H7.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
OX      NCBI_TaxID=562, 83334;
ON      (1)
RN      SEQUENCE FROM N.A.
RA      Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA      Riley M., Collado-Vides J., Glaesner J.D., Rode C.K., Mayhew G.F.,
RA      Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA      Mau B., Shao Y.;
RT      "The complete genome sequence of Escherichia coli K-12.";
RL      Science 277:1453-1474 (1997).
RN      (2)
RN      SEQUENCE FROM N.A.
RC      STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX      MEDLINE=21074935; Pubmed=11206551;
RA      Perna N.T., Plunkett G. III, Burland V., Mau B., Glaesner J.D.,
RA      Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA      Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA      Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA      Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA      Welch R.A., Blattner F.R.;
RT      "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
```

```
Query Match      2.0%; Score 6; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      213 EFFDVL 218
DB      8 EFFDVL 13

RESULT 34
KV3H HUMAN      STANDARD;      PRT;      129 AA.
AC      P04207.
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DE      Hypothetical 14.2 kDa protein.
DT      15-JUL-1999 (Rel. 38, Last annotation update)
DE      Leuconostoc mesenteroides (subsp. cremoris).
OS      Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OC      NCBI_TaxID=33965;
ON      (1)
RN      SEQUENCE FROM N.A.
RA      Cavin J.F., Dartois V.A., Divies C.;
RL      Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC      -1- SIMILARITY: BELONGS TO THE UPF0076 (UK114) FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; M12740; AAA58992.1; -.
DR      PIR; A01898; K3HUC1.
DR      HSSP; P80362; 1WTL.
DR      InterPro; IPR003006; IG_MHC.
DR      InterPro; IPR003596; IG_V.
DR      Pfam; PF00047; Ig; 1.
DR      SMART; SM00406; IGV; 1.
KW      Immunoglobulin V region; Signal.
FT      SIGNAL      1 20
FT      CHAIN      21 129
FT      DOMAIN      21 43
FT      DOMAIN      44 54
FT      DOMAIN      55 69
FT      DOMAIN      70 76
FT      DOMAIN      77 108
FT      DOMAIN      109 118
FT      DOMAIN      119 129
FT      DISULFID      43 108
FT      NON_TER      129 129
SQ      SEQUENCE 129 AA; 14275 MW; 5C13B411BE60CC14 CRC64;

Query Match      2.0%; Score 6; DB 1; Length 129;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      119 GTRVEI 124
DB      122 GTRVEI 127

RESULT 35
Y142 LEUMC
ID      Y142 LEUMC      STANDARD;      PRT;      130 AA.
AC      P91117.
DT      01-NOV-1997 (Rel. 35, Created)
```

RL Nature 409:529-533 (2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RA MEDLINE=21156231; PubMed=11258796;  
RX Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kihara S., Shiba T., Hattori M., Shinagawa H.,  
RA Kihara S., Shiba T., Hattori M., Shinagawa H.,  
RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
RN DNA Res. 8:11-22 (2001).  
RN [4]  
RP CHARACTERIZATION, AND SEQUENCE OF 1-9.  
RX MEDLINE=99085016; PubMed=9867837;  
RA Korbner P., Zander T., Herschlag D., Bardwell J.C.A.;  
RA "A new heat shock protein that binds nucleic acids.";  
RL J. Biol. Chem. 274:249-256 (1999).  
RN [5]  
RP CHARACTERIZATION.  
RX MEDLINE=20141199; PubMed=10675343;  
RA Korbner P., Stahl J.M., Niehaus K.H., Bardwell J.C.A.;  
RT "Hsp15: a ribosome-associated heat shock protein.";  
RL EMBO J. 19:741-748 (2000).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
RX MEDLINE=20141200; PubMed=10675344;  
RA Staker B.L., Korber P., Bardwell J.C.A., Saper M.A.;  
RT "Structure of Hsp15 reveals a novel RNA-binding motif.";  
RL EMBO J. 19:749-757 (2000).  
CC -!- FUNCTION: INVOLVED IN THE RECYCLING OF FREE 50S RIBOSOMAL SUBUNITS  
CC THAT STILL CARRY A NASCENT CHAIN. BINDS RNA MORE SPECIFICALLY THAN  
CC DNA. BINDS WITH VERY HIGH AFFINITY TO THE FREE 50S RIBOSOMAL  
CC SUBUNIT. DOES NOT BIND IT WHEN IT IS PART OF THE 70S RIBOSOME.  
CC -!- SUBUNIT: MONOMER.  
CC -!- INDUCTION: BY HEAT SHOCK.  
CC -!- SIMILARITY: CONTAINS 1 S4 RNA-BINDING DOMAIN.  
CC -!- SIMILARITY: BELONGS TO THE HSP15 FAMILY.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; U18997; AAA58197.1; -.  
DR EMBL; AE000415; AAC76425.1; -.  
DR EMBL; AE005562; AAG58500.1; -.  
DR EMBL; AP002565; BAB37665.1; -.  
DR EcoGene; EGI2929; hslR.  
DR InterPro; IPR002942; S4.  
DR Pfam; PF01479; S4; 1.  
DR SMART; SM00363; S4; 1.  
KW Heat shock; DNA-binding; RNA-binding; Complete proteome.  
FT DOMAIN 9 55  
FT SEQUENCE 133 AA; 15496 MW; 4B3D40FAECC42D5B CRC64;  
Query Match 2.0%; Score 6; DB 1; Length 133;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 149 ALAREM 154  
Db 25 ALAREM 30  
RESULT 37  
IL4\_BOVIN STANDARD; PRT; 135 AA.  
ID IL4\_BOVIN  
AC P30367;  
DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Interleukin-4 precursor (IL-4) (B-cell stimulatory factor 1) (BSF-1)  
DE (Lymphocyte stimulatory factor 1).  
GN IL4.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92290288; PubMed=1601309;  
RA Heussler V.T., Eichhorn M., Dobbelaere D.A.;  
RT "Cloning of a full-length cDNA encoding bovine interleukin 4 by the  
RT polymerase chain reaction.";  
RL Gene 114:273-278 (1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95352962; PubMed=7626886;  
RA Butkamp J., Schwaiger F.W., Solinas-Toldo S., Fries R.,  
RA Eppien J.T.;  
RT "The bovine interleukin-4 gene: genomic organization, localization,  
RT and evolution.";  
RL Mamm. Genome 6:350-356 (1995).  
CC -!- FUNCTION: IL-4 PARTICIPATES IN AT LEAST SEVERAL B-CELL ACTIVATION  
CC PROCESSES AS WELL AS OF OTHER CELL TYPES. IT IS A COSTIMULATOR OF  
CC DNA-SYNTHESIS. IT INDUCES THE EXPRESSION OF CLASS II MHC MOLECULES  
CC ON RESTING B-CELLS. IT ENHANCES BOTH SECRETION AND CELL SURFACE  
CC EXPRESSION OF ICE AND IGGL. IT ALSO REGULATES THE EXPRESSION OF  
CC THE LOW AFFINITY FC RECEPTOR FOR ICE (CD23) ON BOTH LYMPHOCYTES  
CC AND MONOCYTES.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- SIMILARITY: BELONGS TO THE IL-4 / IL-13 FAMILY.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; M71120; AAA30587.1; -.  
DR EMBL; U14160; AAB82730.1; -.  
DR EMBL; U14131; AAB82730.1; JOINED.  
DR EMBL; U14159; AAB82730.1; JOINED.  
DR PIR; JH0684; JH0684.  
DR HSSP; P05112; IILIL.  
DR InterPro; IPR002354; Interleukin\_4.  
DR InterPro; IPR001325; Interleukin\_4\_13.  
DR Pfam; PF00727; IL4; 1.  
DR PRINTS; PR00431; INTERLEUKIN4.  
DR ProDom; PD004070; Interleukin\_4; 1.  
DR SMART; SM00190; IL4\_13; 1.  
DR PROSITE; PS00838; INTERLEUKIN\_4\_13; 1.  
KW Cytokine; Glycoprotein; B-cell activation; Growth factor; Signal.  
FT SIGNAL 1 24  
FT CHAIN 25 135  
FT DISULFID 27 135  
FT DISULFID 48 85  
FT DISULFID 70 105  
FT CARBOHYD 62 62  
FT CONFLICT 5 5  
FT SEQUENCE 135 AA; 15195 MW; E452B7C98654271C CRC64;  
Query Match 2.0%; Score 6; DB 1; Length 135;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 13 VLVCIL 18  
Db 10 VLVCIL 15



```
RESULT 38
IL4_CEREL
ID IL4_CEREL STANDARD; PRT; 135 AA.
AC P51744;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 35, Last annotation update)
DE Interleukin-4 precursor (IL-4) (B-cell stimulatory factor 1) (BSF-1)
DE (Lymphocyte stimulatory factor 1).
GN IL4
OS Cervus elaphus (Red deer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervioidea;
OC Cervidae; Cervinae; Cervus.
OX NCBI_TaxID=9860;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9430479; PubMed=8029646;
RA Hook S.M., Crawford A.M., Chinn D.N., Griffin J.F.T., Buchan G.S.;
RT "Cloning and expression of the cervine interleukin 4 gene.";
RL Scand. J. Immunol. 40:71-76(1994).
CC -1- FUNCTION: IL-4 PARTICIPATES IN AT LEAST SEVERAL B-CELL ACTIVATION
CC PROCESSES AS WELL AS OF OTHER CELL TYPES. IT IS A COSTIMULATOR OF
CC DNA-SYNTHESIS. IT INDUCES THE EXPRESSION OF CLASS II MHC MOLECULES
CC ON RESTING B-CELLS. IT ENHANCES BOTH SECRETION AND CELL SURFACE
CC EXPRESSION OF IGE AND IGG1. IT ALSO REGULATES THE EXPRESSION OF
CC THE LOW AFFINITY FC RECEPTOR FOR IGE (CD23) ON BOTH LYMPHOCYTES
CC AND MONOCYTES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE IL-4 / IL-13 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L07081; AAC37322.1; -.
CC HSP; P05112; ILL.
CC InterPro; IPR002354; Interleukin_4.
CC InterPro; IPR001325; Interleukin_4_13.
CC Pfam; PF00727; IL4; 1.
CC PRINTS; PR00431; INTERLEUKIN4.
CC ProDom; PD004070; Interleukin_4; 1.
CC SMART; SM00190; IL4_13; 1.
CC PROSITE; PS00838; INTERLEUKIN_4_13; 1.
CC Cytokine; Glycoprotein; B-cell activation; Growth factor; Signal.
FT SIGNAL 1 24
FT CHAIN 25 135 INTERLEUKIN-4.
FT DISULFID 27 135 BY SIMILARITY.
FT DISULFID 48 85 BY SIMILARITY.
FT DISULFID 70 105 BY SIMILARITY.
FT CARBOHYD 62 92 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 96 96 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 135 AA; 15156 MW; 6110B6EF31F4F266 CRC64;

Query Match 2.0%; Score 6; DB 1; Length 135;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 13 VLVCLL 18
Db 10 VLVCLL 15

RESULT 39
NUSB_HAEIN
ID NUSB_HAEIN STANDARD; PRT; 144 AA.
AC P45150;
```

```
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE N utilization substance protein B homolog (Nusb protein).
GN NUSB OR H11304.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uutterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: INVOLVED IN THE TRANSCRIPTION TERMINATION PROCESS
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUSB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32810; AAC22951.1; -.
CC HSP; P04381; 1EY1.
CC TIGR; H11304; -.
CC InterPro; IPR000139; Nusb.
CC Pfam; PF01029; Nusb; 1.
CC ProDom; PD005242; Nusb; 1.
CC Transcription termination; RNA-binding; Complete proteome.
KW SEQUENCE 144 AA; 16486 MW; 213A6D820CB9F1CA CRC64;

Query Match 2.0%; Score 6; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 198 SISPYI 203
Db 72 SISPYI 77

RESULT 40
Y127_MYCGE
ID Y127_MYCGE STANDARD; PRT; 145 AA.
AC P47373;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG127.
GN MG127.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
```

RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,  
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,  
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,  
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
RT "The minimal gene complement of Mycoplasma genitalium";  
RL Science 270:397-403(1995).  
CC -!- SIMILARITY: BELONGS TO THE ARSC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U39693; AAC71345.1; -.  
DR TIGR; MG127; -.  
KW Hypothetical protein; Oxidoreductase; Complete proteome.  
SQ SEQUENCE 145 AA; 16565 MW; 1994AEAC391383C2 CRC64;  
Query Match 2.0%; Score 6; DB 1; Length 145;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 211 AIEFFD 216  
DB 28 AIEFFD 33  
|||||

Search completed: July 8, 2003, 10:31:29  
Job time : 46 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 10:29:26 ; Search time 27 Seconds  
(without alignments)  
319.293 Million cell updates/sec

Title: US-10-062-624-40  
Perfect score: 293  
Sequence: 1 MNKLKFTIINTVLVCLLSL.....AVATLNIGYFGGEIGARLTF 293

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PTCUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	293	100.0	293	4	US-09-660-587-40
2	8	2.7	280	4	US-09-660-587-42
3	7	2.4	158	4	US-09-091-725-25
4	7	2.4	274	4	US-09-134-001C-5279
5	7	2.4	283	4	US-09-660-587-4
6	7	2.4	283	4	US-09-261-358A-4
7	7	2.4	409	4	US-09-066-047-19
8	7	2.4	551	2	US-08-067-351-2
9	7	2.4	551	4	US-09-360-490-2
10	7	2.4	1487	3	US-08-840-062-7
11	7	2.4	1529	4	US-09-134-001C-3945
12	6	2.0	30	1	US-08-117-083-25
13	6	2.0	41	4	US-09-143-571-28
14	6	2.0	80	1	US-08-271-354-12
15	6	2.0	80	2	US-08-565-861-12
16	6	2.0	80	5	PCT-US94-07658-12
17	6	2.0	82	1	US-08-497-312-24
18	6	2.0	91	4	US-09-134-001C-5427
19	6	2.0	107	4	US-09-376-330-26
20	6	2.0	111	1	US-07-634-278-47
21	6	2.0	111	1	US-08-477-728-47
22	6	2.0	111	1	US-08-474-040-47
23	6	2.0	111	1	US-08-487-200-47
24	6	2.0	111	4	US-08-484-537-47
25	6	2.0	113	4	US-08-983-607-25
26	6	2.0	128	3	US-09-045-764A-6
27	6	2.0	148	4	US-08-936-165A-481

28	6	2.0	156	4	US-08-936-165A-389	Sequence 389, App
29	6	2.0	158	3	US-09-010-809-18	Sequence 18, Appl
30	6	2.0	159	1	US-08-597-236-9	Sequence 9, Appl
31	6	2.0	159	1	US-08-746-682A-9	Sequence 9, Appl
32	6	2.0	161	4	US-09-615-192A-318	Sequence 318, Appl
33	6	2.0	163	4	US-09-370-838-41	Sequence 41, Appl
34	6	2.0	167	4	US-08-858-207A-353	Sequence 353, App
35	6	2.0	169	4	US-09-149-476-354	Sequence 354, App
36	6	2.0	170	4	US-09-134-001C-3134	Sequence 3134, App
37	6	2.0	173	4	US-09-134-001C-3854	Sequence 3854, App
38	6	2.0	178	4	US-09-495-066-2	Sequence 2, Appl
39	6	2.0	198	4	US-09-134-001C-4050	Sequence 4050, App
40	6	2.0	199	4	US-09-497-779A-6	Sequence 6, Appl
41	6	2.0	207	4	US-08-858-207A-345	Sequence 345, App
42	6	2.0	213	2	US-08-763-121-3	Sequence 3, Appl
43	6	2.0	213	4	US-09-216-066-3	Sequence 3, Appl
44	6	2.0	222	4	US-09-227-357-161	Sequence 161, App
45	6	2.0	231	2	US-08-808-550-33	Sequence 33, Appl
46	6	2.0	240	4	US-09-651-656-1	Sequence 1, Appl
47	6	2.0	240	4	US-09-650-855-1	Sequence 1, Appl
48	6	2.0	250	1	US-08-378-761A-80	Sequence 80, Appl
49	6	2.0	250	1	US-08-485-286-80	Sequence 80, Appl
50	6	2.0	258	4	US-09-497-779A-9	Sequence 9, Appl
51	6	2.0	258	4	US-09-134-001C-5593	Sequence 5593, App
52	6	2.0	263	3	US-08-483-857-5	Sequence 5, Appl
53	6	2.0	263	4	US-09-134-001C-3062	Sequence 3062, App
54	6	2.0	266	4	US-09-814-134A-2	Sequence 2, Appl
55	6	2.0	271	4	US-09-660-587-46	Sequence 46, Appl
56	6	2.0	283	4	US-09-660-587-10	Sequence 10, Appl
57	6	2.0	283	4	US-09-261-358A-10	Sequence 10, Appl
58	6	2.0	283	4	US-09-201-458-6	Sequence 6, Appl
59	6	2.0	291	4	US-09-134-001C-4895	Sequence 4895, App
60	6	2.0	312	4	US-09-134-001C-3458	Sequence 3458, App
61	6	2.0	325	2	US-08-828-922-3	Sequence 3, Appl
62	6	2.0	328	4	US-09-134-001C-3229	Sequence 3229, App
63	6	2.0	329	3	US-09-022-669-2	Sequence 2, Appl
64	6	2.0	329	4	US-09-603-567-2	Sequence 2, Appl
65	6	2.0	330	4	US-09-188-930-148	Sequence 148, App
66	6	2.0	330	4	US-09-188-930-278	Sequence 278, App
67	6	2.0	339	4	US-09-323-873A-2	Sequence 2, Appl
68	6	2.0	342	3	US-08-852-824-2	Sequence 2, Appl
69	6	2.0	344	3	US-08-466-343D-9	Sequence 9, Appl
70	6	2.0	347	1	US-08-461-244-3	Sequence 3, Appl
71	6	2.0	354	4	US-08-724-984A-2	Sequence 2, Appl
72	6	2.0	358	3	US-08-988-876-3	Sequence 3, Appl
73	6	2.0	360	1	US-08-450-393A-4	Sequence 4, Appl
74	6	2.0	360	4	US-08-446-669-4	Sequence 4, Appl
75	6	2.0	360	4	US-09-045-583-50	Sequence 50, Appl
76	6	2.0	360	4	US-09-534-185-50	Sequence 50, Appl
77	6	2.0	360	4	US-08-833-752-7	Sequence 7, Appl
78	6	2.0	360	5	PCT-US95-00476-4	Sequence 4, Appl
79	6	2.0	363	3	US-09-046-086-2	Sequence 2, Appl
80	6	2.0	371	4	US-09-378-088A-64	Sequence 64, Appl
81	6	2.0	374	1	US-08-450-393A-2	Sequence 2, Appl
82	6	2.0	374	4	US-08-446-669-2	Sequence 2, Appl
83	6	2.0	374	5	PCT-US95-00476-2	Sequence 2, Appl
84	6	2.0	384	5	PCT-US92-02091-4	Sequence 4, Appl
85	6	2.0	385	3	US-08-633-993A-11	Sequence 11, Appl
86	6	2.0	385	3	US-08-844-188-11	Sequence 11, Appl
87	6	2.0	385	4	US-09-378-088A-11	Sequence 11, Appl
88	6	2.0	385	4	US-09-378-088A-68	Sequence 68, Appl
89	6	2.0	388	4	US-09-134-001C-3820	Sequence 3820, App
90	6	2.0	391	4	US-09-378-088A-78	Sequence 78, Appl
91	6	2.0	392	4	US-09-378-088A-90	Sequence 90, Appl
92	6	2.0	395	4	US-09-325-932A-182	Sequence 182, App
93	6	2.0	396	6	5349058-2	Patent No. 5349058
94	6	2.0	399	4	US-09-120-772-2	Sequence 2, Appl
95	6	2.0	400	1	US-08-602-010A-8	Sequence 8, Appl
96	6	2.0	400	1	US-08-680-726A-8	Sequence 8, Appl
97	6	2.0	400	4	US-09-092-409-8	Sequence 8, Appl
98	6	2.0	417	4	US-09-134-001C-3810	Sequence 3810, App
99	6	2.0	428	4	US-09-134-001C-2885	Sequence 2885, App
100	6	2.0	448	4	US-09-134-001C-3210	Sequence 3210, App

101	6	2.0	453	2	US-08-851-088-8	Sequence 8, Appli	174	6	2.0	934	3	US-08-709-784-1	Sequence 1, Appli
102	6	2.0	453	4	US-08-851-089-2	Sequence 2, Appli	175	6	2.0	934	4	US-09-651-656-3	Sequence 3, Appli
103	6	2.0	461	4	US-09-134-001C-3316	Sequence 316, Ap	176	6	2.0	934	4	US-09-650-855-3	Sequence 3, Appli
104	6	2.0	462	4	US-09-378-088A-82	Sequence 82, Appl	177	6	2.0	1032	4	US-09-115-954-8	Sequence 8, Appli
105	6	2.0	462	4	US-09-134-001C-4300	Sequence 4300, Ap	178	6	2.0	1044	4	US-09-115-954-2	Sequence 2, Appli
106	6	2.0	466	2	US-08-432-015-4	Sequence 4, Appli	179	6	2.0	1064	3	US-08-726-214-8	Sequence 8, Appli
107	6	2.0	466	2	US-08-684-594-4	Sequence 4, Appli	180	6	2.0	1073	4	US-09-206-942-49	Sequence 49, Appl
108	6	2.0	473	4	US-09-134-001C-3564	Sequence 3564, Ap	181	6	2.0	1077	4	US-09-390-234-12	Sequence 12, Appl
109	6	2.0	474	4	US-09-134-001C-3241	Sequence 3241, Ap	182	6	2.0	1077	4	US-09-412-210-1	Sequence 1, Appli
110	6	2.0	480	2	US-09-171-482-2	Sequence 2, Appli	183	6	2.0	1079	4	US-09-206-942-47	Sequence 47, Appl
111	6	2.0	499	2	US-09-032-315-2	Sequence 2, Appli	184	6	2.0	1154	4	US-09-134-001C-3428	Sequence 3428, Ap
112	6	2.0	499	2	US-08-993-318A-2	Sequence 2, Appli	185	6	2.0	1160	3	US-08-808-599A-24	Sequence 24, Appl
113	6	2.0	499	2	US-09-399-886-2	Sequence 2, Appli	186	6	2.0	1290	1	US-08-470-350B-2	Sequence 2, Appli
114	6	2.0	499	4	US-09-396-260-2	Sequence 2, Appli	187	6	2.0	1463	1	US-08-220-603A-11	Sequence 11, Appl
115	6	2.0	499	4	US-09-576-281-2	Sequence 2, Appli	188	6	2.0	1489	6	5183745-2	Patent No. 5183745
116	6	2.0	500	1	US-08-117-083-68	Sequence 68, Appl	189	6	2.0	1626	2	US-08-771-602D-2	Sequence 2, Appli
117	6	2.0	508	1	US-07-891-942G-10	Sequence 10, Appl	190	6	2.0	1626	4	US-09-232-446B-2	Sequence 2, Appli
118	6	2.0	512	4	US-09-149-922-2	Sequence 2, Appli	191	6	2.0	1705	4	US-08-669-785-4	Sequence 4, Appli
119	6	2.0	514	2	US-08-865-311-2	Sequence 2, Appli	192	6	2.0	1706	4	US-08-669-785-2	Sequence 2, Appli
120	6	2.0	514	4	US-09-315-720-2	Sequence 2, Appli	193	6	2.0	1794	6	5183745-6	Patent No. 5183745
121	6	2.0	516	1	US-08-097-829-4	Sequence 4, Appli	194	6	2.0	1867	2	US-08-479-537A-5	Sequence 5, Appli
122	6	2.0	516	1	US-08-577-403-4	Sequence 4, Appli	195	6	2.0	1867	4	US-09-083-116-5	Sequence 5, Appli
123	6	2.0	520	1	US-08-462-484-2	Sequence 2, Appli	196	6	2.0	1867	4	US-09-134-916A-5	Sequence 5, Appli
124	6	2.0	520	1	US-08-441-147-2	Sequence 2, Appli	197	6	2.0	1872	1	US-08-188-582-14	Sequence 14, Appl
125	6	2.0	520	5	PCT-US95-07536-2	Sequence 2, Appli	198	6	2.0	1872	1	US-08-646-715-14	Sequence 14, Appl
126	6	2.0	529	1	US-07-891-942G-8	Sequence 8, Appli	199	6	2.0	1893	1	US-08-188-582-11	Sequence 11, Appl
127	6	2.0	529	1	US-08-370-909-19	Sequence 19, Appl	200	6	2.0	1893	1	US-08-646-715-11	Sequence 11, Appl
128	6	2.0	534	3	US-09-198-672-2	Sequence 2, Appli	201	6	2.0	2035	2	US-08-479-537A-2	Sequence 2, Appli
129	6	2.0	534	4	US-09-199-290-9	Sequence 9, Appli	202	6	2.0	2035	4	US-09-083-116-2	Sequence 2, Appli
130	6	2.0	534	4	US-09-632-392-2	Sequence 2, Appli	203	6	2.0	2035	4	US-09-134-916A-2	Sequence 2, Appli
131	6	2.0	534	4	US-09-455-679-1	Sequence 1, Appli	204	6	2.0	2165	1	US-08-514-975B-2	Sequence 2, Appli
132	6	2.0	534	4	US-09-351-814-2	Sequence 2, Appli	205	6	2.0	2165	5	PCT-US95-12507-2	Sequence 2, Appli
133	6	2.0	537	4	US-08-540-922D-12	Sequence 12, Appl	206	6	2.0	2227	3	US-08-475-886-2	Sequence 2, Appli
134	6	2.0	546	2	US-09-067-351-1	Sequence 1, Appli	207	6	2.0	2227	3	US-08-475-886-6	Sequence 6, Appli
135	6	2.0	546	4	US-09-360-490-1	Sequence 1, Appli	208	6	2.0	2227	4	US-08-397-232-2	Sequence 2, Appli
136	6	2.0	551	3	US-08-699-103B-25	Sequence 25, Appl	209	6	2.0	2227	4	US-08-397-232-4	Sequence 4, Appli
137	6	2.0	551	4	US-09-229-059-25	Sequence 25, Appl	210	6	2.0	2227	4	US-09-171-387-2	Sequence 2, Appli
138	6	2.0	560	1	US-07-891-942G-5	Sequence 5, Appli	211	6	2.0	2227	4	US-09-653-499-2	Sequence 2, Appli
139	6	2.0	563	4	US-09-134-001C-3172	Sequence 3172, Ap	212	6	2.0	2227	4	US-09-653-499-4	Sequence 4, Appli
140	6	2.0	615	1	US-08-484-106-6	Sequence 6, Appli	213	6	2.0	2227	4	US-09-653-499-6	Sequence 6, Appli
141	6	2.0	615	1	US-08-484-106-6	Sequence 6, Appli	214	6	2.0	2227	4	US-08-469-260A-394	Sequence 394, App
142	6	2.0	624	3	US-08-947-965-78	Sequence 78, Appl	215	6	2.0	2864	4	US-08-894-344C-2	Sequence 2, Appli
143	6	2.0	635	1	US-07-879-617A-9	Sequence 9, Appli	216	6	2.0	2958	2	US-08-477-451-6	Sequence 6, Appli
144	6	2.0	635	1	US-08-753-985-9	Sequence 9, Appli	217	6	2.0	3169	2	US-08-222-617A-13	Sequence 13, Appl
145	6	2.0	640	4	US-09-351-814-13	Sequence 13, Appl	218	6	2.0	3665	2	US-08-222-617A-4	Sequence 4, Appli
146	6	2.0	649	1	US-07-894-212A-8	Sequence 8, Appli	219	6	2.0	3712	2	US-08-222-617A-25	Sequence 25, Appl
147	6	2.0	649	1	US-07-894-212A-2	Sequence 2, Appli	220	6	2.0	3712	2	US-08-222-617A-25	Sequence 25, Appl
148	6	2.0	650	1	US-07-893-928A-1	Sequence 1, Appli	221	6	2.0	4536	4	US-09-180-422B-27	Sequence 27, Appl
149	6	2.0	655	1	US-08-469-203-27	Sequence 27, Appl	222	5	1.7	6	2	US-08-540-412-178	Sequence 178, App
150	6	2.0	655	1	US-08-469-202-28	Sequence 28, Appl	223	5	1.7	6	4	US-09-051-342-178	Sequence 178, App
151	6	2.0	655	2	US-08-484-434C-34	Sequence 34, Appl	224	5	1.7	6	4	US-09-051-342-178	Sequence 178, App
152	6	2.0	655	2	US-08-484-434C-35	Sequence 35, Appl	225	5	1.7	7	4	US-09-258-754-417	Sequence 417, App
153	6	2.0	655	4	US-09-228-986-70	Sequence 70, Appl	226	5	1.7	7	4	US-09-042-107-417	Sequence 417, App
154	6	2.0	658	4	US-09-328-599A-2	Sequence 2, Appli	227	5	1.7	7	4	US-09-461-697-431	Sequence 431, App
155	6	2.0	666	4	US-08-982-785A-11	Sequence 11, Appl	228	5	1.7	8	2	US-08-891-848-11	Sequence 11, Appl
156	6	2.0	668	4	US-09-363-708-4	Sequence 4, Appli	229	5	1.7	8	2	US-08-480-793-8	Sequence 8, Appli
157	6	2.0	676	4	US-09-313-930-2	Sequence 2, Appli	230	5	1.7	8	4	US-08-444-818-592	Sequence 592, App
158	6	2.0	707	4	US-09-134-001C-2962	Sequence 2962, Ap	231	5	1.7	8	5	PCT-US92-05825A-8	Sequence 8, Appli
159	6	2.0	714	4	US-09-115-954-4	Sequence 4, Appli	232	5	1.7	9	1	US-08-080-073-9	Sequence 9, Appli
160	6	2.0	731	2	US-08-313-185-51	Sequence 51, Appl	233	5	1.7	9	2	US-08-475-891A-14	Sequence 14, Appl
161	6	2.0	731	2	US-08-459-499-15	Sequence 15, Appl	234	5	1.7	9	2	US-08-340-283-51	Sequence 51, Appl
162	6	2.0	731	3	US-09-082-614A-51	Sequence 51, Appl	235	5	1.7	9	2	US-08-340-283-150	Sequence 150, App
163	6	2.0	753	3	US-08-674-887A-6	Sequence 6, Appli	236	5	1.7	9	2	US-08-567-375-11	Sequence 11, Appl
164	6	2.0	753	3	US-08-951-844-6	Sequence 6, Appli	237	5	1.7	9	2	US-08-587-680A-18	Sequence 18, Appl
165	6	2.0	810	4	US-09-412-347-6	Sequence 6, Appli	238	5	1.7	9	3	US-08-817-926-14	Sequence 14, Appl
166	6	2.0	810	4	US-09-323-872A-29	Sequence 29, Appl	239	5	1.7	9	4	US-09-142-481-11	Sequence 11, Appl
167	6	2.0	878	4	US-09-556-706B-2	Sequence 2, Appli	240	5	1.7	10	1	US-07-911-531-4	Sequence 4, Appli
168	6	2.0	907	3	US-08-783-774-2	Sequence 2, Appli	241	5	1.7	10	1	US-07-693-636A-4	Sequence 4, Appli
169	6	2.0	907	4	US-09-328-599A-1	Sequence 1, Appli	242	5	1.7	10	1	US-08-237-418-42	Sequence 42, Appl
170	6	2.0	907	5	PCT-US95-04611A-19	Sequence 19, Appl	243	5	1.7	10	1	US-08-424-682A-2	Sequence 2, Appli
171	6	2.0	920	1	US-08-451-715A-19	Sequence 2, Appli	244	5	1.7	10	1	US-08-485-284A-6	Sequence 6, Appli
172	6	2.0	934	1	US-08-457-176-2	Sequence 2, Appli	245	5	1.7	10	3	US-08-159-339A-398	Sequence 398, App
173	6	2.0	934	1	US-08-457-175-2	Sequence 2, Appli	246	5	1.7	10	4	US-08-468-337-42	Sequence 42, Appl

247	5	1.7	10	4	US-09-291-803-13	Sequence 13, Appl	320	5	1.7	20	4	US-09-589-768-30	Sequence 30, Appl
248	5	1.7	10	4	US-09-291-803-14	Sequence 14, Appl	321	5	1.7	20	4	US-09-594-845-30	Sequence 30, Appl
249	5	1.7	10	4	US-09-044-718-41	Sequence 41, Appl	322	5	1.7	20	4	US-09-375-318-41	Sequence 41, Appl
250	5	1.7	11	3	US-08-146-145-14	Sequence 14, Appl	323	5	1.7	20	4	US-09-375-318-42	Sequence 42, Appl
251	5	1.7	11	3	US-08-481-985B-140	Sequence 140, App	324	5	1.7	20	5	PCT-US93-12539-2	Sequence 2, Appl
252	5	1.7	11	4	US-09-053-003-22	Sequence 22, Appl	325	5	1.7	21	4	US-08-469-260A-231	Sequence 231, App
253	5	1.7	11	4	US-08-183-967-20	Sequence 20, Appl	326	5	1.7	22	1	US-08-004-139B-34	Sequence 34, Appl
254	5	1.7	11	6	5177193-19	Patent No. 5177193	327	5	1.7	22	2	US-08-811-492-34	Sequence 34, Appl
255	5	1.7	12	1	US-07-778-233B-15	Sequence 15, Appl	328	5	1.7	22	2	US-08-458-814-10	Sequence 10, Appl
256	5	1.7	12	1	US-07-963-321-15	Sequence 15, Appl	329	5	1.7	22	3	PCT-US96-10545A-34	Sequence 34, Appl
257	5	1.7	12	1	US-08-290-641-15	Sequence 15, Appl	330	5	1.7	23	1	PCT-US96-10545A-34	Sequence 10, Appl
258	5	1.7	12	1	US-08-548-540-15	Sequence 15, Appl	331	5	1.7	23	1	US-08-133-011-14	Sequence 14, Appl
259	5	1.7	12	1	US-08-054-860-9	Sequence 9, Appl	332	5	1.7	23	1	US-08-487-890A-103	Sequence 103, App
260	5	1.7	12	3	US-08-442-378-9	Sequence 9, Appl	333	5	1.7	23	1	US-08-322-730A-14	Sequence 14, Appl
261	5	1.7	12	5	PCT-US96-09809-15	Sequence 15, Appl	334	5	1.7	23	1	US-08-387-874-14	Sequence 14, Appl
262	5	1.7	13	1	US-08-305-871A-27	Sequence 27, Appl	335	5	1.7	23	2	US-08-902-516-15	Sequence 15, Appl
263	5	1.7	14	1	US-07-841-662-27	Sequence 27, Appl	336	5	1.7	23	2	US-08-478-435-103	Sequence 103, App
264	5	1.7	14	1	US-08-209-797-27	Sequence 27, Appl	337	5	1.7	23	2	US-08-337-483-103	Sequence 103, App
265	5	1.7	14	2	US-08-669-685-27	Sequence 27, Appl	338	5	1.7	23	2	US-08-478-373-103	Sequence 103, App
266	5	1.7	14	3	US-09-103-486-27	Sequence 27, Appl	339	5	1.7	23	2	US-08-383-619-14	Sequence 14, Appl
267	5	1.7	14	4	US-09-456-455A-13	Sequence 13, Appl	340	5	1.7	23	3	US-08-474-671-103	Sequence 103, App
268	5	1.7	14	4	US-09-456-455A-12	Sequence 12, Appl	341	5	1.7	23	3	US-08-483-577A-103	Sequence 103, App
269	5	1.7	14	4	US-09-675-922-12	Sequence 12, Appl	342	5	1.7	23	4	US-08-907-739-14	Sequence 14, Appl
270	5	1.7	14	5	PCT-US93-01557-27	Sequence 27, Appl	343	5	1.7	23	4	US-08-897-438-103	Sequence 103, App
271	5	1.7	14	6	5189147-17	Patent No. 5189147	344	5	1.7	23	4	US-08-637-654-103	Sequence 103, App
272	5	1.7	15	2	US-08-408-858A-3	Sequence 3, Appl	345	5	1.7	23	4	US-08-649-518-103	Sequence 103, App
273	5	1.7	15	2	US-08-750-856A-8	Sequence 8, Appl	346	5	1.7	23	4	US-09-449-218D-4	Sequence 4, Appl
274	5	1.7	16	1	US-08-291-349A-3	Sequence 3, Appl	347	5	1.7	23	4	US-09-729-597-14	Sequence 14, Appl
275	5	1.7	16	2	US-09-025-706-8	Sequence 8, Appl	348	5	1.7	23	5	PCT-US93-08364-14	Sequence 14, Appl
276	5	1.7	16	3	US-08-394-748A-11	Sequence 11, Appl	349	5	1.7	24	4	PCT-US96-10905-10	Sequence 10, Appl
277	5	1.7	16	4	US-09-025-622-8	Sequence 8, Appl	350	5	1.7	24	4	US-07-741-453A-16	Sequence 16, Appl
278	5	1.7	16	4	US-09-291-803-3	Sequence 3, Appl	351	5	1.7	24	4	US-09-217-306B-24	Sequence 24, Appl
279	5	1.7	16	5	PCT-US95-02478-11	Sequence 11, Appl	352	5	1.7	25	5	PCT-US96-01600-24	Sequence 24, Appl
280	5	1.7	17	2	US-08-295-643-15	Sequence 15, Appl	353	5	1.7	26	1	US-08-446-692-84	Sequence 84, Appl
281	5	1.7	17	4	US-08-996-679-63	Sequence 63, Appl	354	5	1.7	26	2	US-08-488-351A-84	Sequence 84, Appl
282	5	1.7	17	4	US-08-939-853A-14	Sequence 14, Appl	355	5	1.7	26	4	US-09-135-020-107	Sequence 107, App
283	5	1.7	17	4	US-09-115-335-23	Sequence 23, Appl	356	5	1.7	26	4	US-09-444-871-107	Sequence 107, App
284	5	1.7	17	4	US-09-113-977C-68	Sequence 68, Appl	357	5	1.7	26	4	US-09-597-735-107	Sequence 107, App
285	5	1.7	17	4	US-09-250-059-54	Sequence 54, Appl	358	5	1.7	26	4	US-09-444-295-107	Sequence 107, App
286	5	1.7	17	4	US-09-248-074-50	Sequence 50, Appl	359	5	1.7	26	4	US-09-597-732-107	Sequence 107, App
287	5	1.7	17	4	US-09-187-859-50	Sequence 50, Appl	360	5	1.7	27	2	US-08-637-759B-329	Sequence 329, App
288	5	1.7	17	4	US-09-458-870-54	Sequence 54, Appl	361	5	1.7	27	3	US-08-871-355A-329	Sequence 329, App
289	5	1.7	18	1	US-08-401-512-64	Sequence 64, Appl	362	5	1.7	27	4	US-09-201-945-329	Sequence 329, App
290	5	1.7	18	2	US-07-876-941A-29	Sequence 29, Appl	363	5	1.7	27	4	US-09-149-476-729	Sequence 729, App
291	5	1.7	18	3	US-08-868-594-11	Sequence 11, Appl	364	5	1.7	28	1	US-08-485-692-13	Sequence 13, Appl
292	5	1.7	18	4	US-09-164-186-3	Sequence 3, Appl	365	5	1.7	28	1	US-08-427-072-11	Sequence 11, Appl
293	5	1.7	18	4	US-09-164-186-12	Sequence 12, Appl	366	5	1.7	28	1	US-08-419-519-13	Sequence 13, Appl
294	5	1.7	18	4	US-09-236-012-99	Sequence 99, Appl	367	5	1.7	28	4	US-08-318-288-1	Sequence 1, Appl
295	5	1.7	18	4	US-09-461-697-429	Sequence 429, App	368	5	1.7	28	4	US-09-282-357-1	Sequence 1, Appl
296	5	1.7	18	4	US-09-434-323-11	Sequence 11, Appl	369	5	1.7	28	4	US-08-900-230-50	Sequence 50, Appl
297	5	1.7	18	4	US-08-637-670-31	Sequence 31, Appl	370	5	1.7	28	4	US-09-604-871-3	Sequence 3, Appl
298	5	1.7	18	6	5171684-19	Patent No. 5171684	371	5	1.7	28	5	PCT-US96-01720-5	Sequence 5, Appl
299	5	1.7	19	4	US-08-975-040-10	Sequence 10, Appl	372	5	1.7	28	5	PCT-US96-01720-6	Sequence 6, Appl
300	5	1.7	20	1	US-08-237-418-44	Sequence 44, Appl	373	5	1.7	29	2	US-08-310-912A-17	Sequence 17, Appl
301	5	1.7	20	1	US-08-479-236-2	Sequence 2, Appl	374	5	1.7	29	2	US-08-455-079-1	Sequence 1, Appl
302	5	1.7	20	2	US-08-598-873-53	Sequence 53, Appl	375	5	1.7	29	3	US-09-045-632-84	Sequence 84, Appl
303	5	1.7	20	2	US-08-934-915-33	Sequence 33, Appl	376	5	1.7	29	3	US-08-841-089-17	Sequence 17, Appl
304	5	1.7	20	2	US-08-934-915-138	Sequence 138, App	377	5	1.7	29	4	US-09-301-085-17	Sequence 17, Appl
305	5	1.7	20	2	US-08-993-976-2	Sequence 2, Appl	378	5	1.7	29	4	US-09-302-812-12	Sequence 12, Appl
306	5	1.7	20	3	US-08-300-928C-29	Sequence 29, Appl	379	5	1.7	29	4	US-09-511-477-12	Sequence 12, Appl
307	5	1.7	20	3	US-08-430-944D-29	Sequence 29, Appl	380	5	1.7	29	4	US-09-511-507-12	Sequence 12, Appl
308	5	1.7	20	3	US-08-430-014-29	Sequence 29, Appl	381	5	1.7	29	5	PCT-US95-04570-17	Sequence 17, Appl
309	5	1.7	20	3	US-08-467-023-51	Sequence 51, Appl	382	5	1.7	29	5	PCT-US95-04589-17	Sequence 17, Appl
310	5	1.7	20	3	US-08-467-023-52	Sequence 52, Appl	383	5	1.7	30	1	US-08-407-831-1	Sequence 1, Appl
311	5	1.7	20	3	US-08-695-301A-30	Sequence 30, Appl	384	5	1.7	30	1	US-08-300-928C-31	Sequence 31, Appl
312	5	1.7	20	3	US-08-893-526A-30	Sequence 30, Appl	385	5	1.7	30	3	US-08-430-944D-31	Sequence 31, Appl
313	5	1.7	20	3	US-08-695-304C-30	Sequence 30, Appl	386	5	1.7	30	3	US-08-430-014-31	Sequence 31, Appl
314	5	1.7	20	3	US-08-431-184-29	Sequence 29, Appl	387	5	1.7	30	3	US-08-431-184-31	Sequence 31, Appl
315	5	1.7	20	4	US-08-612-973-71	Sequence 71, Appl	388	5	1.7	30	4	US-08-905-223-356	Sequence 356, App
316	5	1.7	20	4	US-08-468-337-44	Sequence 44, Appl	389	5	1.7	30	4	US-09-262-856A-2	Sequence 2, Appl
317	5	1.7	20	4	US-08-605-430-53	Sequence 53, Appl	390	5	1.7	30	5	PCT-US94-04361-28	Sequence 28, Appl
318	5	1.7	20	4	US-09-927-597-71	Sequence 71, Appl	391	5	1.7	31	1	US-08-053-131-91	Sequence 91, Appl
319	5	1.7	20	4	US-09-199-290-2	Sequence 2, Appl	392	5	1.7	31	1	US-08-645-641-91	Sequence 91, Appl

393	5	1.7	31	1	US-07-853-408B-91	*Sequence 91, Appl	466	1.7	35	4	US-09-315-304B-451	Sequence 451, App
394	5	1.7	31	2	US-08-096-762-91	Sequence 91, Appl	467	1.7	35	4	US-09-315-304B-452	Sequence 452, App
395	5	1.7	31	2	US-08-308-865-91	Sequence 91, Appl	468	1.7	35	4	US-09-315-304B-453	Sequence 453, App
396	5	1.7	31	4	US-09-042-353-288	Sequence 288, App	469	1.7	35	4	US-09-315-304B-454	Sequence 454, App
397	5	1.7	31	4	US-08-758-417A-136	Sequence 136, App	470	1.7	35	4	US-09-315-304B-455	Sequence 455, App
398	5	1.7	31	5	PCT-US92-10983-91	Sequence 91, Appl	471	1.7	35	4	US-09-315-304B-456	Sequence 456, App
399	5	1.7	32	1	US-08-526-583-5	Sequence 5, Appl	472	1.7	36	4	US-09-215-212-8	Sequence 8, Appl
400	5	1.7	32	2	US-08-526-583-6	Sequence 6, Appl	473	1.7	36	6	5451527-5	Patent No. 5451527
401	5	1.7	32	2	US-08-526-583-7	Sequence 7, Appl	474	1.7	36	6	5451527-6	Patent No. 5451527
402	5	1.7	32	2	US-08-526-583-8	Sequence 8, Appl	475	1.7	37	2	US-08-743-200-10	Sequence 10, Appl
403	5	1.7	32	4	US-09-227-357-223	Sequence 223, App	476	1.7	37	2	US-08-709-924-25	Sequence 25, Appl
404	5	1.7	32	4	US-09-120-653D-28	Sequence 28, App	477	1.7	37	2	US-08-709-925-25	Sequence 25, Appl
405	5	1.7	34	6	5451527-4	Patent No. 5451527	478	1.7	37	3	US-08-486-099-207	Sequence 207, App
406	5	1.7	35	2	US-08-491-527A-2	Sequence 2, Appl	479	1.7	37	3	US-08-484-223B-207	Sequence 207, App
407	5	1.7	35	2	US-08-845-659-2	Sequence 2, Appl	480	1.7	37	3	US-08-919-597-207	Sequence 207, App
408	5	1.7	35	3	US-08-486-099-204	Sequence 204, App	481	1.7	37	3	US-08-475-668A-207	Sequence 207, App
409	5	1.7	35	3	US-08-486-099-206	Sequence 206, App	482	1.7	37	3	US-08-485-551A-207	Sequence 207, App
410	5	1.7	35	3	US-08-484-223B-204	Sequence 204, App	483	1.7	37	3	US-08-471-913A-207	Sequence 207, App
411	5	1.7	35	3	US-08-484-223B-206	Sequence 206, App	484	1.7	37	4	US-08-905-223-321	Sequence 321, App
412	5	1.7	35	3	US-08-919-597-204	Sequence 204, App	485	1.7	37	4	US-08-485-264A-207	Sequence 207, App
413	5	1.7	35	3	US-08-919-597-206	Sequence 206, App	486	1.7	37	4	US-08-709-948-25	Sequence 25, Appl
414	5	1.7	35	3	US-08-475-668A-204	Sequence 204, App	487	1.7	37	4	US-08-474-349A-207	Sequence 207, App
415	5	1.7	35	3	US-08-475-668A-206	Sequence 206, App	488	1.7	38	2	US-08-902-516-47	Sequence 47, Appl
416	5	1.7	35	3	US-08-485-551A-204	Sequence 204, App	489	1.7	39	1	US-08-006-676B-3	Sequence 3, Appl
417	5	1.7	35	3	US-08-485-551A-206	Sequence 206, App	490	1.7	39	1	US-08-282-845-3	Sequence 3, Appl
418	5	1.7	35	3	US-08-471-913A-204	Sequence 204, App	491	1.7	39	2	US-08-657-392-29	Sequence 29, Appl
419	5	1.7	35	3	US-08-471-913A-206	Sequence 206, App	492	1.7	39	2	US-08-428-414A-5	Sequence 5, Appl
420	5	1.7	35	3	US-08-845-658-2	Sequence 2, Appl	493	1.7	39	3	US-08-851-843A-23	Sequence 23, Appl
421	5	1.7	35	4	US-08-485-264A-204	Sequence 204, App	494	1.7	39	3	US-08-974-549A-213	Sequence 213, App
422	5	1.7	35	4	US-08-485-264A-206	Sequence 206, App	495	1.7	39	4	US-08-854-050-23	Sequence 23, Appl
423	5	1.7	35	4	US-09-082-279B-441	Sequence 441, App	496	1.7	39	4	US-09-430-323-23	Sequence 23, Appl
424	5	1.7	35	4	US-09-082-279B-442	Sequence 442, App	497	1.7	39	5	PCT-US94-00324-3	Sequence 3, Appl
425	5	1.7	35	4	US-09-082-279B-443	Sequence 443, App	498	1.7	39	5	PCT-US94-02539-29	Sequence 29, Appl
426	5	1.7	35	4	US-09-082-279B-444	Sequence 444, App	499	1.7	39	6	5451527-8	Patent No. 5451527
427	5	1.7	35	4	US-09-082-279B-445	Sequence 445, App	500	1.7	39	6	5451527-12	Patent No. 5451527
428	5	1.7	35	4	US-09-082-279B-446	Sequence 446, App	501	1.7	40	1	US-07-977-630-76	Sequence 76, Appl
429	5	1.7	35	4	US-09-082-279B-447	Sequence 447, App	502	1.7	40	4	US-08-908-371B-10	Sequence 10, Appl
430	5	1.7	35	4	US-09-082-279B-448	Sequence 448, App	503	1.7	40	4	US-09-106-568E-119	Sequence 119, App
431	5	1.7	35	4	US-09-082-279B-449	Sequence 449, App	504	1.7	41	6	5451527-9	Patent No. 5451527
432	5	1.7	35	4	US-09-082-279B-450	Sequence 450, App	505	1.7	41	6	5451527-10	Patent No. 5451527
433	5	1.7	35	4	US-09-082-279B-451	Sequence 451, App	506	1.7	41	6	5451527-11	Patent No. 5451527
434	5	1.7	35	4	US-09-082-279B-452	Sequence 452, App	507	1.7	41	6	5451527-13	Patent No. 5451527
435	5	1.7	35	4	US-09-082-279B-453	Sequence 453, App	508	1.7	41	6	5451527-14	Patent No. 5451527
436	5	1.7	35	4	US-09-082-279B-454	Sequence 454, App	509	1.7	42	1	US-08-239-256-17	Sequence 17, Appl
437	5	1.7	35	4	US-09-082-279B-455	Sequence 455, App	510	1.7	42	4	US-07-757-022B-24	Sequence 24, Appl
438	5	1.7	35	4	US-09-082-279B-456	Sequence 456, App	511	1.7	42	5	PCT-US93-08739-3	Sequence 3, Appl
439	5	1.7	35	4	US-09-518-178-2	Sequence 2, Appl	512	1.7	43	1	US-07-998-003A-71	Sequence 71, Appl
440	5	1.7	35	4	US-08-474-349A-204	Sequence 204, App	513	1.7	43	1	US-08-453-274B-71	Sequence 71, Appl
441	5	1.7	35	4	US-08-474-349A-206	Sequence 206, App	514	1.7	43	1	US-08-453-695A-71	Sequence 71, Appl
442	5	1.7	35	4	US-08-474-349A-470	Sequence 470, App	515	1.7	43	1	US-08-268-161A-71	Sequence 42, Appl
443	5	1.7	35	4	US-08-474-349A-471	Sequence 471, App	516	1.7	43	2	US-08-488-161-42	Sequence 42, Appl
444	5	1.7	35	4	US-08-474-349A-472	Sequence 472, App	517	1.7	43	2	US-08-453-702A-71	Sequence 71, Appl
445	5	1.7	35	4	US-08-474-349A-473	Sequence 473, App	518	1.7	43	3	US-09-273-685-42	Sequence 42, Appl
446	5	1.7	35	4	US-08-474-349A-474	Sequence 474, App	519	1.7	43	4	US-09-099-639-71	Sequence 71, Appl
447	5	1.7	35	4	US-08-474-349A-475	Sequence 475, App	520	1.7	43	4	US-08-908-371B-12	Sequence 12, Appl
448	5	1.7	35	4	US-08-474-349A-476	Sequence 476, App	521	1.7	43	5	PCT-US93-12588-71	Sequence 71, Appl
449	5	1.7	35	4	US-08-474-349A-477	Sequence 477, App	522	1.7	43	5	PCT-US95-08071-71	Sequence 71, Appl
450	5	1.7	35	4	US-08-474-349A-478	Sequence 478, App	523	1.7	43	5	PCT-US95-11534-42	Sequence 42, Appl
451	5	1.7	35	4	US-08-474-349A-479	Sequence 479, App	524	1.7	44	1	US-08-361-920-15	Sequence 15, Appl
452	5	1.7	35	4	US-08-474-349A-480	Sequence 480, App	525	1.7	44	1	US-08-479-939-15	Sequence 15, Appl
453	5	1.7	35	4	US-08-474-349A-481	Sequence 481, App	526	1.7	44	1	US-08-483-432-15	Sequence 15, Appl
454	5	1.7	35	4	US-08-474-349A-482	Sequence 482, App	527	1.7	45	1	US-08-262-037-88	Sequence 88, Appl
455	5	1.7	35	4	US-08-474-349A-483	Sequence 483, App	528	1.7	45	2	US-08-637-759B-102	Sequence 102, App
456	5	1.7	35	4	US-09-315-304B-441	Sequence 441, App	529	1.7	45	3	US-08-486-099-205	Sequence 205, App
457	5	1.7	35	4	US-09-315-304B-442	Sequence 442, App	530	1.7	45	3	US-08-871-355A-102	Sequence 102, App
458	5	1.7	35	4	US-09-315-304B-443	Sequence 443, App	531	1.7	45	3	US-08-484-223B-205	Sequence 205, App
459	5	1.7	35	4	US-09-315-304B-444	Sequence 444, App	532	1.7	45	3	US-08-919-597-205	Sequence 205, App
460	5	1.7	35	4	US-09-315-304B-445	Sequence 445, App	533	1.7	45	3	US-08-475-668A-205	Sequence 205, App
461	5	1.7	35	4	US-09-315-304B-446	Sequence 446, App	534	1.7	45	3	US-08-485-551A-205	Sequence 205, App
462	5	1.7	35	4	US-09-315-304B-447	Sequence 447, App	535	1.7	45	3	US-08-471-913A-205	Sequence 205, App
463	5	1.7	35	4	US-09-315-304B-448	Sequence 448, App	536	1.7	45	4	US-09-060-756-733	Sequence 733, App
464	5	1.7	35	4	US-09-315-304B-449	Sequence 449, App	537	1.7	45	4	US-08-485-264A-205	Sequence 205, App
465	5	1.7	35	4	US-09-315-304B-450	Sequence 450, App	538	1.7	45	4	US-08-900-230-13	Sequence 13, Appl

539	5	1.7	45	4	US-08-474-349A-205	Sequence 205, App	612	5	1.7	75	4	US-09-134-001C-5348	Sequence 5348, Ap
540	5	1.7	45	4	US-09-201-945-102	Sequence 102, App	613	5	1.7	76	2	US-08-465-380-29	Sequence 29, Appl
541	5	1.7	45	6	5451527-1	Patent No. 5451527	614	5	1.7	76	2	US-08-480-478-58	Sequence 58, Appl
542	5	1.7	46	1	US-08-118-270-199	Sequence 199, App	615	5	1.7	76	2	US-08-486-397-29	Sequence 29, Appl
543	5	1.7	46	5	PCT-US93-08528-199	Sequence 199, App	616	5	1.7	76	2	US-08-486-399-29	Sequence 29, Appl
544	5	1.7	47	4	US-09-082-279B-783	Sequence 783, App	617	5	1.7	76	2	US-08-461-965-29	Sequence 29, Appl
545	5	1.7	47	4	US-09-315-304B-783	Sequence 783, App	618	5	1.7	76	2	US-08-326-110A-58	Sequence 58, Appl
546	5	1.7	48	3	US-08-804-439A-101	Sequence 101, App	619	5	1.7	76	2	US-08-634-641-29	Sequence 29, Appl
547	5	1.7	48	3	US-08-405-647B-42	Sequence 42, Appl	620	5	1.7	76	2	US-09-249-471-29	Sequence 29, Appl
548	5	1.7	48	4	US-08-985-499-42	Sequence 42, Appl	621	5	1.7	76	3	US-09-249-472-29	Sequence 29, Appl
549	5	1.7	48	5	PCT-US96-03180-42	Sequence 42, Appl	622	5	1.7	76	3	US-09-249-451-29	Sequence 29, Appl
550	5	1.7	49	4	US-09-082-279B-782	Sequence 782, App	623	5	1.7	76	3	US-08-809-455-29	Sequence 29, Appl
551	5	1.7	49	4	US-09-315-304B-782	Sequence 782, App	624	5	1.7	76	3	US-09-249-461-29	Sequence 29, Appl
552	5	1.7	50	1	US-08-200-016-5	Sequence 5, Appl	625	5	1.7	76	3	US-09-249-468-29	Sequence 29, Appl
553	5	1.7	50	1	US-08-178-477B-16	Sequence 16, Appl	626	5	1.7	76	3	US-08-468-674B-73	Sequence 73, Appl
554	5	1.7	52	4	US-08-988-856B-14	Sequence 14, Appl	627	5	1.7	78	1	US-08-780-571-73	Sequence 73, Appl
555	5	1.7	54	1	US-08-468-674B-55	Sequence 55, Appl	628	5	1.7	78	3	US-08-946-026-26	Sequence 26, Appl
556	5	1.7	54	1	US-08-780-571-55	Sequence 55, Appl	629	5	1.7	78	3	US-08-102-757-17	Sequence 17, Appl
557	5	1.7	55	1	US-08-262-037-99	Sequence 89, Appl	630	5	1.7	80	1	US-08-225-757B-11	Sequence 11, Appl
558	5	1.7	55	1	US-08-262-037-118	Sequence 118, App	631	5	1.7	80	1	US-08-225-757B-14	Sequence 14, Appl
559	5	1.7	56	1	US-08-468-674B-57	Sequence 57, Appl	632	5	1.7	80	1	US-08-233-597-4	Sequence 4, Appl
560	5	1.7	56	1	US-08-468-674B-59	Sequence 59, Appl	633	5	1.7	80	5	PCT-US94-01202-4	Sequence 4, Appl
561	5	1.7	56	1	US-08-780-571-57	Sequence 57, Appl	634	5	1.7	81	1	US-08-225-757B-8	Sequence 8, Appl
562	5	1.7	56	1	US-08-780-571-59	Sequence 59, Appl	635	5	1.7	81	2	US-08-974-565C-3	Sequence 3, Appl
563	5	1.7	58	1	US-08-468-674B-63	Sequence 63, Appl	636	5	1.7	81	3	US-09-255-748-3	Sequence 3, Appl
564	5	1.7	58	1	US-08-446-646-13	Sequence 13, Appl	637	5	1.7	83	1	US-08-279-590A-6	Sequence 6, Appl
565	5	1.7	58	1	US-08-780-571-63	Sequence 63, Appl	638	5	1.7	83	2	US-08-910-092-6	Sequence 6, Appl
566	5	1.7	58	2	US-08-465-380-60	Sequence 60, Appl	639	5	1.7	83	3	US-09-100-804-24	Sequence 24, Appl
567	5	1.7	58	2	US-08-486-397-60	Sequence 60, Appl	640	5	1.7	83	3	US-08-851-843A-9	Sequence 9, Appl
568	5	1.7	58	2	US-08-486-399-60	Sequence 60, Appl	641	5	1.7	83	4	US-08-974-549A-191	Sequence 191, App
569	5	1.7	58	2	US-08-461-965-60	Sequence 60, Appl	642	5	1.7	83	4	US-08-854-050-9	Sequence 9, Appl
570	5	1.7	58	2	US-08-634-641-60	Sequence 60, Appl	643	5	1.7	83	4	US-09-430-323-9	Sequence 9, Appl
571	5	1.7	58	3	US-09-249-471-60	Sequence 60, Appl	644	5	1.7	84	1	US-08-279-590A-4	Sequence 4, Appl
572	5	1.7	58	3	US-09-249-472-60	Sequence 60, Appl	645	5	1.7	84	2	US-08-910-092-4	Sequence 4, Appl
573	5	1.7	58	3	US-09-249-451-60	Sequence 60, Appl	646	5	1.7	84	2	US-08-353-476-79	Sequence 79, Appl
574	5	1.7	58	3	US-08-809-455-60	Sequence 60, Appl	647	5	1.7	84	4	US-08-679-493A-98	Sequence 98, Appl
575	5	1.7	58	3	US-09-249-461-60	Sequence 60, Appl	648	5	1.7	84	4	US-09-134-001C-2919	Sequence 2919, Ap
576	5	1.7	58	3	US-09-249-448-60	Sequence 60, Appl	649	5	1.7	85	1	US-08-341-219-22	Sequence 22, Appl
577	5	1.7	58	4	US-09-164-186-8	Sequence 8, Appl	650	5	1.7	85	3	US-08-851-843A-11	Sequence 11, Appl
578	5	1.7	59	1	US-08-470-179-11	Sequence 11, Appl	651	5	1.7	85	4	US-08-974-549A-193	Sequence 193, App
579	5	1.7	59	1	US-08-446-692-109	Sequence 109, App	652	5	1.7	85	4	US-08-912-114A-22	Sequence 22, Appl
580	5	1.7	59	2	US-08-488-351A-109	Sequence 109, App	653	5	1.7	85	4	US-08-854-050-11	Sequence 11, Appl
581	5	1.7	60	1	US-08-262-037-90	Sequence 90, Appl	654	5	1.7	85	4	US-09-430-323-11	Sequence 11, Appl
582	5	1.7	60	2	US-08-607-412-2	Sequence 2, Appl	655	5	1.7	87	1	US-07-867-194-6	Sequence 6, Appl
583	5	1.7	60	4	US-09-291-803-1	Sequence 1, Appl	656	5	1.7	88	1	US-08-203-676-21	Sequence 21, Appl
584	5	1.7	60	4	US-09-227-357-518	Sequence 518, App	657	5	1.7	88	1	US-08-468-674B-75	Sequence 75, Appl
585	5	1.7	62	4	US-08-630-915A-135	Sequence 135, App	658	5	1.7	88	1	US-08-780-571-75	Sequence 75, Appl
586	5	1.7	64	1	US-08-091-569-22	Sequence 22, Appl	659	5	1.7	88	2	US-08-327-362-2	Sequence 2, Appl
587	5	1.7	64	1	US-08-203-676-22	Sequence 22, Appl	660	5	1.7	88	2	US-08-338-530A-4	Sequence 4, Appl
588	5	1.7	64	2	US-08-822-238-22	Sequence 22, App	661	5	1.7	88	2	US-08-709-924-24	Sequence 24, Appl
589	5	1.7	64	4	US-09-134-001C-3006	Sequence 3006, Ap	662	5	1.7	88	2	US-08-822-238-21	Sequence 21, Appl
590	5	1.7	65	4	US-09-615-192A-334	Sequence 334, App	663	5	1.7	88	2	US-08-709-925-24	Sequence 24, Appl
591	5	1.7	65	4	US-08-857-076-61	Sequence 61, Appl	664	5	1.7	88	2	US-09-158-565-2	Sequence 2, Appl
592	5	1.7	66	4	US-08-857-076-90	Sequence 90, Appl	665	5	1.7	88	4	US-08-444-818-83	Sequence 83, Appl
593	5	1.7	69	4	US-09-134-001C-3082	Sequence 3082, Ap	666	5	1.7	88	4	US-09-267-384-4	Sequence 4, Appl
594	5	1.7	69	4	US-09-134-001C-3613	Sequence 3613, Ap	667	5	1.7	88	4	US-08-709-948-24	Sequence 24, Appl
595	5	1.7	71	4	US-08-446-137B-5	Sequence 5, Appl	668	5	1.7	89	1	US-08-468-674B-41	Sequence 41, Appl
596	5	1.7	72	1	US-08-370-225-31	Sequence 31, Appl	669	5	1.7	89	1	US-08-780-571-41	Sequence 41, Appl
597	5	1.7	72	1	US-08-468-674B-61	Sequence 61, Appl	670	5	1.7	89	1	US-08-467-023-64	Sequence 64, Appl
598	5	1.7	72	1	US-08-461-859-31	Sequence 31, Appl	671	5	1.7	90	3	US-08-468-674B-45	Sequence 45, Appl
599	5	1.7	72	1	US-08-482-828B-6	Sequence 6, Appl	672	5	1.7	91	1	US-08-780-571-45	Sequence 45, Appl
600	5	1.7	72	1	US-08-780-571-61	Sequence 61, Appl	673	5	1.7	91	1	US-08-905-223-421	Sequence 421, App
601	5	1.7	72	1	US-08-486-036A-6	Sequence 6, Appl	674	5	1.7	91	4	US-08-300-928C-13	Sequence 13, Appl
602	5	1.7	72	4	US-09-005-298-6	Sequence 6, Appl	675	5	1.7	92	3	US-08-430-944D-13	Sequence 13, Appl
603	5	1.7	72	4	US-08-768-619-6	Sequence 6, Appl	676	5	1.7	92	3	US-08-430-014-13	Sequence 13, Appl
604	5	1.7	72	4	US-08-489-260A-233	Sequence 233, App	677	5	1.7	92	3	US-08-431-184-13	Sequence 13, Appl
605	5	1.7	72	5	PCT-US93-10069-31	Sequence 31, Appl	678	5	1.7	92	6	5284931-8	Patent No. 5284931
606	5	1.7	72	5	PCT-US96-09848-6	Sequence 6, Appl	679	5	1.7	93	4	US-08-905-223-402	Sequence 402, App
607	5	1.7	74	4	US-09-227-357-265	Sequence 265, App	680	5	1.7	93	4	US-09-025-596-92	Sequence 92, Appl
608	5	1.7	74	4	US-09-227-357-631	Sequence 631, App	681	5	1.7	93	4	US-08-858-207A-521	Sequence 521, App
609	5	1.7	74	4	US-09-149-476-408	Sequence 408, App	682	5	1.7	93	4	US-08-974-549A-164	Sequence 164, App
610	5	1.7	75	1	US-08-624-663A-4	Sequence 4, Appl	683	5	1.7	94	4	US-09-157-210-12	Sequence 12, Appl
611	5	1.7	75	4	US-09-134-001C-4078	Sequence 4078, Ap	684	5	1.7	94	4		

685	5	1.7	94	4	US-09-149-476-505	Sequence 505, App	758	5	1.7	108	1	US-08-360-125-12	Sequence 12, Appl
686	5	1.7	95	1	US-08-338-009B-2	Sequence 2, Appli	759	5	1.7	108	2	US-08-450-578-12	Sequence 12, Appl
687	5	1.7	95	1	US-08-597-583-2	Sequence 2, Appli	760	5	1.7	108	2	US-08-650-262-2	Sequence 2, Appli
688	5	1.7	95	2	US-09-047-125-30	Sequence 30, Appl	761	5	1.7	108	2	US-09-017-628-12	Sequence 12, Appl
689	5	1.7	95	3	US-08-946-339A-74	Sequence 74, Appl	762	5	1.7	108	3	US-09-014-880-12	Sequence 12, Appl
690	5	1.7	95	3	US-07-736-335E-30	Sequence 30, Appl	763	5	1.7	108	3	US-08-868-594-2	Sequence 2, Appli
691	5	1.7	95	5	PCT-US95-14717-2	Sequence 2, Appli	764	5	1.7	108	4	US-09-240-374-66	Sequence 66, Appl
692	5	1.7	95	1	US-08-480-449-24	Sequence 24, Appl	765	5	1.7	108	4	US-09-025-769B-20	Sequence 20, Appl
693	5	1.7	96	2	US-08-660-542-24	Sequence 24, Appl	766	5	1.7	108	4	US-08-434-323-2	Sequence 2, Appli
694	5	1.7	96	4	US-08-679-493A-162	Sequence 162, App	767	5	1.7	108	4	US-08-450-363-12	Sequence 12, Appl
695	5	1.7	96	4	US-08-479-603-24	Sequence 24, Appl	768	5	1.7	109	1	US-07-807-529A-6	Sequence 6, Appli
696	5	1.7	96	4	US-09-134-001C-4685	Sequence 4685, Ap	769	5	1.7	109	1	US-08-478-039-91	Sequence 91, Appl
697	5	1.7	97	1	US-08-211-202-127	Sequence 127, App	770	5	1.7	109	1	US-07-987-272A-8	Sequence 8, Appli
698	5	1.7	97	1	US-08-211-202-129	Sequence 129, App	771	5	1.7	109	1	US-08-476-349A-91	Sequence 91, Appl
699	5	1.7	97	3	US-08-946-026-29	Sequence 29, Appl	772	5	1.7	109	4	US-09-202-181-2	Sequence 2, Appli
700	5	1.7	97	3	US-09-023-221A-8	Sequence 8, Appli	773	5	1.7	109	4	US-09-157-370-5	Sequence 5, Appli
701	5	1.7	97	8	US-09-282-352A-8	Sequence 8, Appli	774	5	1.7	109	4	US-09-370-838-44	Sequence 44, Appl
702	5	1.7	98	2	US-08-709-924-23	Sequence 23, Appl	775	5	1.7	109	5	PCT-US93-02462-6	Sequence 6, Appli
703	5	1.7	98	2	US-08-709-925-23	Sequence 23, Appl	776	5	1.7	110	1	US-08-466-886-29	Sequence 29, Appl
704	5	1.7	98	3	US-08-881-037-75	Sequence 75, Appl	777	5	1.7	110	4	US-08-469-617-29	Sequence 29, Appl
705	5	1.7	98	4	US-08-709-948-23	Sequence 23, Appl	778	5	1.7	111	1	US-07-682-193-4	Sequence 4, Appli
706	5	1.7	99	4	US-09-227-357-630	Sequence 630, App	779	5	1.7	111	1	US-07-634-278-54	Sequence 54, Appl
707	5	1.7	100	1	US-08-488-113B-154	Sequence 154, App	780	5	1.7	111	1	US-08-477-728-54	Sequence 54, Appl
708	5	1.7	100	1	US-08-477-484B-154	Sequence 154, App	781	5	1.7	111	1	US-08-338-009B-4	Sequence 4, Appli
709	5	1.7	100	1	US-08-107-669D-18	Sequence 18, Appl	782	5	1.7	111	1	US-08-597-583-4	Sequence 4, Appli
710	5	1.7	100	1	US-08-472-788A-18	Sequence 18, Appl	783	5	1.7	111	1	US-08-442-542-4	Sequence 4, Appli
711	5	1.7	100	2	US-08-477-531B-18	Sequence 18, Appl	784	5	1.7	111	1	US-08-474-040-54	Sequence 54, Appl
712	5	1.7	100	2	US-08-646-360-154	Sequence 154, App	785	5	1.7	111	1	US-08-487-200-54	Sequence 54, Appl
713	5	1.7	100	2	US-08-082-842A-18	Sequence 18, Appl	786	5	1.7	111	3	US-08-300-928C-6	Sequence 6, Appli
714	5	1.7	100	4	US-08-839-765-154	Sequence 154, App	787	5	1.7	111	3	US-08-430-944D-6	Sequence 6, Appli
715	5	1.7	100	4	US-09-136-389-154	Sequence 154, App	788	5	1.7	111	3	US-08-430-014-6	Sequence 6, Appli
716	5	1.7	100	4	US-09-610-838-154	Sequence 154, App	789	5	1.7	111	3	US-08-483-749A-22	Sequence 22, Appl
717	5	1.7	101	2	US-08-627-873-5	Sequence 5, Appli	790	5	1.7	111	3	US-08-765-469-4	Sequence 4, Appli
718	5	1.7	101	3	US-08-866-381A-4	Sequence 4, Appli	791	5	1.7	111	3	US-08-881-037-76	Sequence 76, Appl
719	5	1.7	101	3	US-08-894-173-89	Sequence 89, Appl	792	5	1.7	111	3	US-08-431-184-6	Sequence 6, Appli
720	5	1.7	101	4	US-09-398-193-89	Sequence 89, Appl	793	5	1.7	111	4	US-08-484-537-54	Sequence 54, Appl
721	5	1.7	101	4	US-08-469-260A-59	Sequence 59, Appl	794	5	1.7	111	4	US-09-149-476-340	Sequence 340, App
722	5	1.7	102	1	US-08-400-256-36	Sequence 36, Appl	795	5	1.7	111	5	PCT-US95-14717-4	Sequence 4, Appli
723	5	1.7	102	1	US-08-463-090B-8	Sequence 8, Appli	796	5	1.7	112	4	US-08-980-523-10	Sequence 10, Appl
724	5	1.7	102	2	US-08-750-856A-14	Sequence 14, Appl	797	5	1.7	112	4	US-09-134-001C-2950	Sequence 2950, Ap
725	5	1.7	102	2	US-08-750-856A-15	Sequence 15, Appl	798	5	1.7	112	6	5171684-4	Patent No. 5171684
726	5	1.7	102	2	US-08-750-856A-16	Sequence 16, Appl	799	5	1.7	113	1	US-08-211-202-115	Sequence 115, App
727	5	1.7	102	2	US-08-750-856A-17	Sequence 17, Appl	800	5	1.7	113	1	US-08-494-577-3	Sequence 3, Appli
728	5	1.7	102	2	US-08-750-856A-18	Sequence 18, Appl	801	5	1.7	113	1	US-08-494-577-8	Sequence 8, Appli
729	5	1.7	102	2	US-08-750-856A-19	Sequence 19, Appl	802	5	1.7	113	2	US-08-553-457A-6	Sequence 6, Appli
730	5	1.7	102	2	US-08-710-749-8	Sequence 8, Appli	803	5	1.7	113	2	US-08-553-457A-10	Sequence 10, Appl
731	5	1.7	102	3	US-08-975-365-36	Sequence 36, Appl	804	5	1.7	113	2	US-08-795-868-8	Sequence 3, Appli
732	5	1.7	102	3	US-09-023-221A-2	Sequence 2, Appli	805	5	1.7	113	2	US-08-795-868-18	Sequence 18, Appl
733	5	1.7	102	4	US-09-282-352A-2	Sequence 2, Appli	806	5	1.7	113	3	US-09-023-221A-20	Sequence 20, Appl
734	5	1.7	102	4	US-09-134-001C-5181	Sequence 5181, Ap	807	5	1.7	113	3	US-09-282-352A-20	Sequence 20, Appl
735	5	1.7	103	3	US-08-881-037-40	Sequence 40, Appl	808	5	1.7	113	4	US-09-303-069-3	Sequence 3, Appli
736	5	1.7	104	2	US-08-598-873-7	Sequence 7, Appli	809	5	1.7	113	4	US-09-303-069-4	Sequence 4, Appli
737	5	1.7	104	3	US-08-894-173-72	Sequence 72, Appl	810	5	1.7	113	4	US-09-303-069-8	Sequence 8, Appli
738	5	1.7	104	4	US-09-398-193-72	Sequence 72, Appl	811	5	1.7	113	4	US-09-303-069-18	Sequence 18, Appl
739	5	1.7	104	4	US-08-605-430-7	Sequence 7, Appli	812	5	1.7	113	4	US-09-134-250-3	Sequence 3, Appli
740	5	1.7	104	4	US-09-615-192A-135	Sequence 35, App	813	5	1.7	113	4	US-09-134-250-4	Sequence 4, Appli
741	5	1.7	105	1	US-08-488-113B-157	Sequence 157, App	814	5	1.7	113	4	US-09-134-250-8	Sequence 8, Appli
742	5	1.7	105	1	US-08-477-484B-157	Sequence 157, App	815	5	1.7	113	4	US-09-134-250-18	Sequence 18, Appl
743	5	1.7	105	1	US-08-107-669D-21	Sequence 21, Appl	816	5	1.7	113	4	US-08-414-926A-14	Sequence 14, Appl
744	5	1.7	105	1	US-08-472-788A-21	Sequence 21, Appl	817	5	1.7	114	1	US-08-385-241-3	Sequence 3, Appli
745	5	1.7	105	2	US-08-477-531B-21	Sequence 21, Appl	818	5	1.7	114	1	US-08-926-922-14	Sequence 14, Appl
746	5	1.7	105	2	US-08-646-360-157	Sequence 157, App	819	5	1.7	114	2	US-09-253-682-14	Sequence 14, Appl
747	5	1.7	105	2	US-08-082-842A-21	Sequence 21, App	820	5	1.7	114	3	US-09-347-819-4	Sequence 4, Appli
748	5	1.7	105	4	US-08-839-765-157	Sequence 157, App	821	5	1.7	114	4	US-09-527-657-14	Sequence 14, Appl
749	5	1.7	105	4	US-09-136-389-157	Sequence 157, App	822	5	1.7	114	4	US-09-127-946-6	Sequence 6, Appli
750	5	1.7	105	4	US-09-610-838-157	Sequence 157, App	823	5	1.7	114	4	US-08-483-749A-24	Sequence 24, Appl
751	5	1.7	105	4	US-09-134-001C-5273	Sequence 5273, Ap	824	5	1.7	115	3	US-09-267-177-32	Sequence 32, Appl
752	5	1.7	106	4	US-09-010-233-4	Sequence 4, Appli	825	5	1.7	115	4	US-09-012-669F-37	Sequence 37, Appl
753	5	1.7	106	6	5204258-4	Patent No. 5204258	826	5	1.7	117	4	US-08-936-165A-347	Patent No. 5242821
754	5	1.7	107	4	US-09-025-769B-34	Sequence 34, Appl	827	5	1.7	117	4	5242821-12	Sequence 17, Appl
755	5	1.7	107	4	US-09-025-769B-55	Sequence 55, Appl	828	5	1.7	117	6	US-09-413-814-17	Sequence 17, Appl
756	5	1.7	108	1	US-08-259-372A-10	Sequence 10, Appl	829	5	1.7	118	4	US-08-685-871-66	Sequence 66, Appl
757	5	1.7	108	1	US-08-468-671-10	Sequence 10, Appl	830	5	1.7	119	3		



831	5	1.7	119	4	US-09-406-532-2	Sequence 2, Appli	904	5	1.7	144	2	US-08-786-606-1	Sequence 1, Appli
832	5	1.7	120	1	US-08-111-080-25	Sequence 25, Appli	905	5	1.7	144	2	US-08-786-606-7	Sequence 7, Appli
833	5	1.7	120	1	US-08-211-202-1	Sequence 1, Appli	906	5	1.7	144	2	US-08-824-405-13	Sequence 13, Appli
834	5	1.7	120	1	US-08-211-980-25	Sequence 25, Appli	907	5	1.7	144	4	US-09-046-604-20	Sequence 20, Appli
835	5	1.7	120	4	US-08-679-006-34	Sequence 34, Appli	908	5	1.7	144	4	US-09-134-001C-4218	Sequence 4218, Ap
836	5	1.7	120	4	US-08-543-246B-18	Sequence 18, Appli	909	5	1.7	145	1	US-08-425-673-1	Sequence 1, Appli
837	5	1.7	120	4	US-09-071-035-206	Sequence 206, App	910	5	1.7	145	1	US-08-425-673-2	Sequence 2, Appli
838	5	1.7	120	5	PCT-US93-07967-25	Sequence 25, Appli	911	5	1.7	145	1	US-08-425-673-10	Sequence 10, Appli
839	5	1.7	122	1	US-07-634-278-88	Sequence 88, Appli	912	5	1.7	145	1	US-08-298-189B-1	Sequence 1, Appli
840	5	1.7	122	1	US-08-477-728-88	Sequence 88, Appli	913	5	1.7	145	1	US-08-475-213-10	Sequence 10, Appli
841	5	1.7	122	1	US-08-474-040-88	Sequence 88, Appli	914	5	1.7	145	2	US-08-395-238-2	Sequence 2, Appli
842	5	1.7	122	1	US-08-487-200-88	Sequence 88, Appli	915	5	1.7	145	2	US-08-640-977-6	Sequence 6, Appli
843	5	1.7	122	3	US-08-987-144-2	Sequence 2, Appli	916	5	1.7	145	4	US-09-142-320-4	Sequence 4, Appli
844	5	1.7	122	4	US-08-484-537-88	Sequence 88, Appli	917	5	1.7	145	4	US-09-142-320-11	Sequence 11, Appli
845	5	1.7	124	1	US-08-307-499-2	Sequence 2, Appli	918	5	1.7	145	4	US-09-142-320-12	Sequence 12, Appli
846	5	1.7	124	1	US-08-446-646-3	Sequence 3, Appli	919	5	1.7	145	4	US-09-142-320-13	Sequence 13, Appli
847	5	1.7	124	4	US-09-012-669F-36	Sequence 36, Appli	920	5	1.7	145	4	US-09-142-320-14	Sequence 14, Appli
848	5	1.7	124	1	US-09-299-268-2	Sequence 2, Appli	921	5	1.7	145	4	US-09-142-320-15	Sequence 15, Appli
849	5	1.7	124	1	US-08-462-949-24	Sequence 24, Appli	922	5	1.7	145	4	US-09-142-320-16	Sequence 16, Appli
850	5	1.7	125	1	US-08-023-764B-24	Sequence 24, Appli	923	5	1.7	145	4	US-08-918-288-68	Sequence 68, Appli
851	5	1.7	125	4	US-09-134-001C-3462	Sequence 3462, Ap	924	5	1.7	145	4	US-09-282-357-68	Sequence 68, Appli
852	5	1.7	126	4	US-09-342-647-16	Sequence 16, Appli	925	5	1.7	145	4	US-08-908-371B-1	Sequence 1, Appli
853	5	1.7	126	4	US-09-134-001C-3766	Sequence 3766, Ap	926	5	1.7	146	3	US-08-476-120-8	Sequence 8, Appli
854	5	1.7	127	1	US-08-446-646-7	Sequence 7, Appli	927	5	1.7	147	4	US-08-965-762-30	Sequence 30, Appli
855	5	1.7	128	4	US-08-859-207A-314	Sequence 314, App	928	5	1.7	147	4	US-09-911-882-30	Sequence 30, Appli
856	5	1.7	128	4	US-08-134-001C-3090	Sequence 3090, Ap	929	5	1.7	147	4	US-09-911-882-30	Sequence 30, Appli
857	5	1.7	130	4	US-08-444-818-81	Sequence 81, Appli	930	5	1.7	148	1	US-07-708-038-4	Sequence 4, Appli
858	5	1.7	131	1	US-07-634-278-67	Sequence 67, Appli	931	5	1.7	148	1	US-08-127-995-4	Sequence 4, Appli
859	5	1.7	131	1	US-08-477-728-67	Sequence 67, Appli	932	5	1.7	149	2	US-08-874-460-2	Sequence 2, Appli
860	5	1.7	131	1	US-08-474-040-67	Sequence 67, Appli	933	5	1.7	151	2	US-08-637-759B-349	Sequence 349, App
861	5	1.7	131	1	US-08-487-200-67	Sequence 67, Appli	934	5	1.7	151	2	US-08-367-942C-42	Sequence 42, Appli
862	5	1.7	131	3	US-08-836-561-25	Sequence 25, Appli	935	5	1.7	151	3	US-08-871-355A-349	Sequence 349, App
863	5	1.7	131	4	US-08-484-537-67	Sequence 67, Appli	936	5	1.7	151	4	US-09-201-945-349	Sequence 349, App
864	5	1.7	131	4	US-09-247-155-175	Sequence 175, App	937	5	1.7	152	1	US-08-450-945-71	Sequence 71, Appli
865	5	1.7	132	1	US-08-111-939-18	Sequence 18, Appli	938	5	1.7	152	4	US-08-976-161-71	Sequence 71, Appli
866	5	1.7	133	4	US-08-953-326-20	Sequence 20, Appli	939	5	1.7	152	4	US-09-120-653D-7	Sequence 7, Appli
867	5	1.7	133	4	US-09-660-587-7	Sequence 7, Appli	940	5	1.7	152	4	US-09-071-035-128	Sequence 128, App
868	5	1.7	133	4	US-0261-358A-7	Sequence 7, Appli	941	5	1.7	153	2	US-08-791-924-1	Sequence 1, Appli
869	5	1.7	133	4	US-09-201-458-3	Sequence 3, Appli	942	5	1.7	153	3	US-08-943-173-1	Sequence 1, Appli
870	5	1.7	134	4	US-09-199-637A-243	Sequence 243, App	943	5	1.7	153	4	US-09-160-246-15	Sequence 15, Appli
871	5	1.7	135	4	US-09-091-117-1	Sequence 1, Appli	944	5	1.7	153	4	US-08-936-165A-507	Sequence 507, App
872	5	1.7	135	2	US-08-543-246B-17	Sequence 17, Appli	945	5	1.7	154	1	US-08-153-848-36	Sequence 36, Appli
873	5	1.7	136	4	US-08-477-451-31	Sequence 31, Appli	946	5	1.7	154	3	US-09-299-843A-36	Sequence 36, Appli
874	5	1.7	136	4	US-09-171-461-5	Sequence 5, Appli	947	5	1.7	154	4	US-09-088-337B-36	Sequence 36, Appli
875	5	1.7	136	4	US-09-183-861-49	Sequence 49, Appli	948	5	1.7	154	4	US-08-887-534A-78	Sequence 78, Appli
876	5	1.7	136	4	US-09-022-765-49	Sequence 49, Appli	949	5	1.7	154	5	PCT-US93-11153-36	Sequence 36, Appli
877	5	1.7	136	4	US-09-134-001C-4819	Sequence 4819, Ap	950	5	1.7	155	2	US-08-801-972-1	Sequence 1, Appli
878	5	1.7	137	2	US-08-933-750C-49	Sequence 49, Appli	951	5	1.7	155	3	US-09-378-881-1	Sequence 1, Appli
879	5	1.7	137	3	US-09-125-265-2	Sequence 2, Appli	952	5	1.7	157	1	US-07-956-700B-34	Sequence 34, Appli
880	5	1.7	137	4	US-09-234-613-49	Sequence 49, Appli	953	5	1.7	157	1	US-08-476-537-34	Sequence 34, Appli
881	5	1.7	137	4	US-09-254-722-1	Sequence 1, Appli	954	5	1.7	157	1	US-08-485-607-34	Sequence 34, Appli
882	5	1.7	137	4	US-09-381-737-1	Sequence 1, Appli	955	5	1.7	157	2	US-08-475-879-34	Sequence 34, Appli
883	5	1.7	137	4	US-09-149-476-393	Sequence 393, App	956	5	1.7	157	2	US-08-883-070-3	Sequence 3, Appli
884	5	1.7	137	4	US-09-269-019-1	Sequence 1, Appli	957	5	1.7	157	4	US-09-433-043B-34	Sequence 34, Appli
885	5	1.7	139	4	US-08-444-818-171	Sequence 171, App	958	5	1.7	158	4	US-08-849-634B-2	Sequence 2, Appli
886	5	1.7	139	4	US-08-444-818-172	Sequence 172, App	959	5	1.7	158	4	US-09-134-001C-5407	Sequence 5407, Ap
887	5	1.7	139	4	US-08-444-818-174	Sequence 174, App	960	5	1.7	159	3	US-08-481-435-13	Sequence 13, Appli
888	5	1.7	139	4	US-08-444-818-175	Sequence 175, App	961	5	1.7	159	4	US-09-430-503-34	Sequence 34, Appli
889	5	1.7	139	4	PCT-US93-03076-9	Sequence 5124, Ap	962	5	1.7	159	4	US-09-430-503-36	Sequence 36, Appli
890	5	1.7	140	5	US-08-336-618-18	Sequence 18, Appli	963	5	1.7	159	4	US-09-430-503-38	Sequence 38, Appli
891	5	1.7	142	1	US-08-387-942C-41	Sequence 41, Appli	964	5	1.7	159	4	US-08-586-272-3	Sequence 3, Appli
892	5	1.7	142	2	US-08-945-983-7	Sequence 7, Appli	965	5	1.7	160	3	US-09-082-969-3	Sequence 3, Appli
893	5	1.7	142	5	PCT-US92-03993-7	Sequence 7, Appli	966	5	1.7	160	3	US-07-961-702-2	Sequence 2, Appli
894	5	1.7	143	2	US-08-345-321-8	Sequence 8, Appli	967	5	1.7	162	1	US-08-472-284-2	Sequence 2, Appli
895	5	1.7	143	3	US-09-012-084-2	Sequence 2, Appli	968	5	1.7	162	1	US-08-850-117-4	Sequence 4, Appli
896	5	1.7	143	3	US-08-969-683A-61	Sequence 61, Appli	969	5	1.7	162	1	US-09-008-181-4	Sequence 4, Appli
897	5	1.7	143	4	US-09-250-609-17	Sequence 17, Appli	970	5	1.7	162	1	US-08-472-418-2	Sequence 2, Appli
898	5	1.7	144	1	US-07-874-848B-2	Sequence 2, Appli	971	5	1.7	162	2	US-09-068-140A-2	Sequence 2, Appli
899	5	1.7	144	1	US-08-440-103-20	Sequence 20, Appli	972	5	1.7	162	2	US-09-562-737-71	Sequence 71, Appli
900	5	1.7	144	1	US-08-440-542-20	Sequence 20, Appli	973	5	1.7	163	4	US-09-562-737-75	Sequence 75, Appli
901	5	1.7	144	1	US-08-231-368-20	Sequence 20, Appli	974	5	1.7	163	4	US-09-562-737-76	Sequence 76, Appli
902	5	1.7	144	1	US-08-440-210-20	Sequence 20, Appli	975	5	1.7	163	4		
903	5	1.7	144	1	US-08-440-210-20	Sequence 20, Appli	976	5	1.7	163	4		

977 5 1.7 163 4 US-09-562-737-77 Sequence 77, Appl  
978 5 1.7 163 4 US-09-562-737-78 Sequence 78, Appl  
979 5 1.7 163 4 US-09-562-737-79 Sequence 79, Appl  
980 5 1.7 164 4 US-09-522-433B-20 Sequence 20, Appl  
981 5 1.7 164 4 US-09-522-433B-21 Sequence 21, Appl  
982 5 1.7 165 2 US-08-709-924-2 Sequence 2, Appl  
983 5 1.7 165 2 US-08-709-925-2 Sequence 2, Appl  
984 5 1.7 165 4 US-08-709-948-2 Sequence 2, Appl  
985 5 1.7 165 4 US-08-858-207A-462 Sequence 462, App  
986 5 1.7 165 4 US-08-978-283-2 Sequence 2, Appl  
987 5 1.7 166 2 US-08-918-723-4 Sequence 4, Appl  
988 5 1.7 166 2 US-09-237-507-4 Sequence 4, Appl  
989 5 1.7 167 4 US-09-166-350-13 Sequence 13, Appl  
990 5 1.7 170 4 US-09-370-473-2 Sequence 2, Appl  
991 5 1.7 170 4 US-09-311-626B-10 Sequence 10, Appl  
992 5 1.7 170 4 US-09-487-792-13 Sequence 13, Appl  
993 5 1.7 171 2 US-08-934-959-2 Sequence 2, Appl  
994 5 1.7 173 4 US-09-068-140A-13 Sequence 13, Appl  
995 5 1.7 173 4 US-09-232-191-35 Sequence 35, Appl  
996 5 1.7 173 4 US-09-232-200-87 Sequence 87, Appl  
997 5 1.7 173 4 US-09-232-197-87 Sequence 87, Appl  
998 5 1.7 173 4 US-09-232-201-87 Sequence 87, Appl  
999 5 1.7 173 4 US-09-134-001C-4275 Sequence 4275, Ap  
1000 5 1.7 174 1 US-08-460-806-19 Sequence 19, Appl

## ALIGNMENTS

RESULT 1  
US-09-660-587-40  
; Sequence 40, Application US/09660587  
; Patent No. 6392023  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: 09/660,587  
; CURRENT FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 40  
; LENGTH: 293  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of E. canis p28-1 protein  
US-09-660-587-40

Query Match 100.0%; Score 293; DB 4; Length 293;  
Best Local Similarity 100.0%; Pred. No. 1.8e-273; Mismatches 0; Indels 0; Gaps 0;  
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MNKLKFTIINTVLVCLLSLPNISSSKAINNAKYYGLYISQYKPSVSVFNSFKET 60  
DB 1 MNKLKFTIINTVLVCLLSLPNISSSKAINNAKYYGLYISQYKPSVSVFNSFKET 60  
QY 61 NVITKNLTKKDVDSIEFTKTDASVGSINPSNFTPIPTAVFDQNSVNFNGTIGYTFAGT 120  
DB 61 NVITKNLTKKDVDSIEFTKTDASVGSINPSNFTPIPTAVFDQNSVNFNGTIGYTFAGT 120  
QY 121 RVEIGSEVEFDVKNPGGYTLSDAYRYFALAREMKNSTPKKYSNSIFHTVWMDGLS 180  
DB 121 RVEIGSEVEFDVKNPGGYTLSDAYRYFALAREMKNSTPKKYSNSIFHTVWMDGLS 180  
QY 181 IISVIWVCYDPSLNNLSISPIYCGAGVDAIEFFDLVHLKPAYOSKLGIAVSLPSNISL 240  
DB 181 IISVIWVCYDPSLNNLSISPIYCGAGVDAIEFFDLVHLKPAYOSKLGIAVSLPSNISL 240

QY 241 FASLYYHKVMGNQPNLVQHVLAELASIPKITSAVATNLNIGYFGGEIGARLTF 293  
DB 241 FASLYYHKVMGNQPNLVQHVLAELASIPKITSAVATNLNIGYFGGEIGARLTF 293  
RESULT 2  
US-09-660-587-42  
; Sequence 42, Application US/09660587  
; Patent No. 6392023  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: 09/660,587  
; CURRENT FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 42  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of E. canis p28-2 protein  
US-09-660-587-42

Query Match 2.7%; Score 8; DB 4; Length 280;  
Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 281 GYFGGEIG 288  
DB 268 GYFGGEIG 275  
RESULT 3  
US-09-091-725-25  
; Sequence 25, Application US/09091725  
; Patent No. 6329141  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Improved methods for transforming Phaffia  
; TITLE OF INVENTION: and recombinant DNA for use therein  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morrison & Foerster llp  
; STREET: 2000 Pennsylvania Avenue, N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: United States of America  
; ZIP: 20006-1888  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/091,725  
; FILING DATE: 23-DEC-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 95203620.0  
; FILING DATE: 22-DEC-1995  
; APPLICATION NUMBER: EP 96200943.7  
; FILING DATE: 11-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: E. Victor Donahue  
; REGISTRATION NUMBER: 35,492  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:

;  
; LENGTH: 158 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-091-725-25

Query Match 2.4%; Score 7; DB 4; Length 158;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 114 YTFAGT 120  
Db 148 YTFAGT 154

## RESULT 4

US-09-134-001C-5279  
; Sequence 5279, Application US/09134001C  
; Patent No. 6380370

## GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5279

; LENGTH: 274

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-5279

Query Match 2.4%; Score 7; DB 4; Length 274;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 117 AEGTRVE 123  
Db 50 AEGTRVE 56

## RESULT 5

US-09-660-587-4  
; Sequence 4, Application US/09660587  
; Patent No. 6392023

## GENERAL INFORMATION:

; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: US/09/660,587  
; CURRENT FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 4

; LENGTH: 283

; TYPE: PRT

; ORGANISM: Ehrlichia canis

; FEATURE:

; OTHER INFORMATION: amino acid sequence of p28-5 protein

US-09-660-587-4

Query Match 2.4%; Score 7; DB 4; Length 283;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 FDKVKNPG 137  
Db 125 FDKVKNPG 131

## RESULT 6

US-09-261-358A-4  
; Sequence 4, Application US/09261358A  
; Patent No. 6403780

## GENERAL INFORMATION:

; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP  
; CURRENT APPLICATION NUMBER: US/09/261,358A  
; CURRENT FILING DATE: 1999-03-03  
; PRIOR APPLICATION NUMBER: 09/201,458  
; PRIOR FILING DATE: 1998-11-30  
; NUMBER OF SEQ ID NOS: 33  
; SEQ ID NO 4

; LENGTH: 283

; TYPE: PRT

; ORGANISM: Ehrlichia canis

; FEATURE:

; OTHER INFORMATION: amino acid sequence of Eca28SA2 protein

US-09-261-358A-4

Query Match 2.4%; Score 7; DB 4; Length 283;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 FDKVKNPG 137  
Db 125 FDKVKNPG 131

## RESULT 7

US-09-066-047-19  
; Sequence 19, Application US/09066047A  
; Patent No. 6306394

## GENERAL INFORMATION:

; APPLICANT: MURPHY, Cheryl  
; APPLICANT: STOREY, James  
; APPLICANT: BELTZ, Gerald A.  
; COUGHLIN, Richard T.  
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND METHODS OF  
; USE OF GRANULOCYTIC ERLICHIA  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HALE AND DORR LLP  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: United States  
; ZIP: 02109

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/066,047A  
; FILING DATE: 24-Apr-1998  
; CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/044,869  
; FILING DATE: 25-APR-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Superko, Colleen  
; REGISTRATION NUMBER: 39,850

REFERENCE/DOCKET NUMBER: 106.941.156  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 526-6000  
TELEFAX: (617) 526-5000  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 409 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 19:  
US-09-066-047-19

Query Match 2.4%; Score 7; DB 4; Length 409;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 282 YGGEIG 288  
DB 398 YGGEIG 404

RESULT 8  
US-09-067-351-2  
Sequence 2, Application US/09067351  
Patent No. 5994081  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Corley, Neil C.  
APPLICANT: Baughn, Mariah  
TITLE OF INVENTION: HUMAN KERATINS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/067,351  
FILING DATE: Herewith  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: CERRONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0511 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 551 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: KERANOT02  
CLONE: 2029060

Query Match 2.4%; Score 7; DB 2; Length 551;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 ALKKDVD 75

US-09-067-351-2  
Query Match 2.4%; Score 7; DB 2; Length 551;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 ALKKDVD 75

Db 255 ALKKDVD 261

RESULT 9  
US-09-360-490-2  
Sequence 2, Application US/09360490  
Patent No. 6221843  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Corley, Neil C.  
APPLICANT: Baughn, Mariah  
TITLE OF INVENTION: HUMAN KERATINS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/360,490  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/067,351  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: CERRONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0511 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 551 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: KERANOT02  
CLONE: 2029060

Query Match 2.4%; Score 7; DB 4; Length 551;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 ALKKDVD 75

Db 255 ALKKDVD 261

RESULT 10  
US-08-840-062-7  
Sequence 7, Application US/08840062  
Patent No. 611977  
GENERAL INFORMATION:  
APPLICANT: LASKY, LAURENCE A.  
APPLICANT: WU, XAI  
TITLE OF INVENTION: TYPE C LECTINS  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd

; CITY: South San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94080  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: WinPatIn (Genentech)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/840,062  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Dreger, Girger R.  
 ; REGISTRATION NUMBER: 33,055  
 ; REFERENCE/DOCKET NUMBER: P1019R1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415/225-3216  
 ; TELEFAX: 415/952-9861  
 ; TELEX: 910/371-7168  
 ; INFORMATION FOR SEQ ID NO: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1487 amino acids  
 ; TYPE: Amino acid  
 ; TOPOLOGY: Linear  
 ; US-08-840-062-7

Query Match 2.4%; Score 7; DB 3; Length 1487;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 263 AELASIP 269  
 Db 1146 AELASIP 1152

## RESULT 11

US-09-134-001C-3945  
 ; Sequence 3945, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 3945  
 ; LENGTH: 1529  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 ; US-09-134-001C-3945

Query Match 2.4%; Score 7; DB 4; Length 1529;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 AYQSKLG 229  
 Db 241 AYQSKLG 247

## RESULT 12

US-08-117-083-25  
 ; Sequence 25, Application US/08117083  
 ; Patent No. 5719054  
 ; GENERAL INFORMATION:

; APPLICANT: Boursnell, Michael E.  
 ; APPLICANT: Inglis, Stephen C.  
 ; APPLICANT: Munro, Alan J.  
 ; TITLE OF INVENTION: Recombinant Virus Vectors Encoding Human  
 ; TITLE OF INVENTION: Papilloma Virus Proteins  
 ; NUMBER OF SEQUENCES: 70  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Walter H. Dreger  
 ; STREET: 4 Embarcadero Center, Suite 3400  
 ; CITY: San Francisco  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94111  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/117,083  
 ; FILING DATE: 10-SEP-1993  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Dreger, Walter H.  
 ; REGISTRATION NUMBER: 24,190  
 ; REFERENCE/DOCKET NUMBER: A-58783  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415-781-1989  
 ; TELEFAX: 415-398-3249  
 ; TELEX: 910 277299  
 ; INFORMATION FOR SEQ ID NO: 25:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 30 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-117-083-25

Query Match 2.0%; Score 6; DB 1; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 85;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 51 VFSNFS 56  
 Db 23 VFSNFS 28

## RESULT 13

US-09-143-571-28  
 ; Sequence 28, Application US/09143571  
 ; Patent No. 633153  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FISHEL, Richard  
 ; APPLICANT: GRADIA, Scott  
 ; APPLICANT: ACHARYA, Samir  
 ; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR EFFECTING ADENINE  
 ; TITLE OF INVENTION: NUCLEOTIDE MODULATION OF DNA MISMATCH RECOGNITION  
 ; TITLE OF INVENTION: PROTEINS  
 ; FILE REFERENCE: 9855-6U1  
 ; CURRENT APPLICATION NUMBER: US/09/143,571  
 ; CURRENT FILING DATE: 1998-08-28  
 ; EARLIER APPLICATION NUMBER: 60/093,935  
 ; EARLIER FILING DATE: 1998-07-23  
 ; EARLIER APPLICATION NUMBER: 60/066,977  
 ; EARLIER FILING DATE: 1997-11-28  
 ; EARLIER APPLICATION NUMBER: 60/057,136  
 ; EARLIER FILING DATE: 1997-08-28  
 ; NUMBER OF SEQ ID NOS: 36  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 28  
 ; LENGTH: 41  
 ; TYPE: PRT

```
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: region V of hMSH2
US-09-143-571-28

Query Match          2.0%; Score 6; DB 4; Length 41;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      261 HVAELA 266
Db      19 HVAELA 24

RESULT 14
US-08-271-354-12
; Sequence 12, Application US/08271354
; Patent No. 5695983
; GENERAL INFORMATION:
; APPLICANT: Miller, Samuel I.
; APPLICANT: Mekalanos, John J.
; TITLE OF INVENTION: SALMONELLA VACCINES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/271,354
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/271,354
; FILING DATE: 06-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/090,526
; FILING DATE: 09-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/629,602
; FILING DATE: 18-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/303001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-271-354-12

Query Match          2.0%; Score 6; DB 1; Length 80;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      223 AYQSKL 228
Db      46 AYQSKL 51

RESULT 15
US-08-565-861-12
; Sequence 12, Application US/08565861
```

```
; Patent No. 5843426
; GENERAL INFORMATION:
; APPLICANT: Miller, Samuel I.
; APPLICANT: Mekalanos, John J.
; APPLICANT: Hohmann, Elizabeth
; TITLE OF INVENTION: SALMONELLA VACCINES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/565,861
; FILING DATE: 01-DEC-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/271,354
; FILING DATE: 06-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/090,526
; FILING DATE: 09-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/629,602
; FILING DATE: 18-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/303001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-565-861-12

Query Match          2.0%; Score 6; DB 2; Length 80;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      223 AYQSKL 228
Db      46 AYQSKL 51

RESULT 16
PCT-US94-07658-12
; Sequence 12, Application PC/TUS9407658
; GENERAL INFORMATION:
; APPLICANT: Miller, Samuel I.
; APPLICANT: Mekalanos, John J.
; TITLE OF INVENTION: SALMONELLA VACCINES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version
; SOFTWARE: #1.25 and Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07658
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/090,526
; FILING DATE: July 9, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/220001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US94-07658-12

Query Match 2.0%; Score 6; DB 5; Length 80;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 AYQSKL 228
Db 46 AYQSKL 51

RESULT 17
US-08-497-312-24
; Sequence 24, Application US/08497312
; Patent No. 5712120
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Method for obtaining modified
; TITLE OF INVENTION: immunoglobulins with reduced immunogenicity of murine
; TITLE OF INVENTION: antibody variable domains, compositions containing them.
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CENTRO DE INMUNOLOGIA MOLECULAR
; STREET: 215 Y 15, ATABEY PLAYA
; CITY: HAVANA
; STATE:
; COUNTRY: CUBA
; ZIP: 11600
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA: US/08/497,312
; FILING DATE: 30-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CU 80/94
; FILING DATE: 30-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BOND, LAURENCE B.
; REGISTRATION NUMBER: 30,549
; REFERENCE/DOCKET NUMBER: 2629US
; TELEPHONE: 801/531-9168
; TELEFAX: 801/531-9168
; TELEX: 388961 1PM04UT
```

```
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 82 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-497-312-24
```

```
Query Match 2.0%; Score 6; DB 1; Length 82;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 119 GTRVEI 124
Db 74 GTRVEI 79
```

## RESULT 18

```
US-09-134-001C-5427
; Sequence 5427, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5427
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-5427
```

```
Query Match 2.0%; Score 6; DB 4; Length 91;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 180 SIISVI 185
Db 51 SIISVI 56
```

## RESULT 19

```
US-09-376-330-26
; Sequence 26, Application US/09376330
; Patent No. 6399321
; GENERAL INFORMATION:
; APPLICANT: Tessier, Daniel C.
; APPLICANT: Dignard, Daniel
; APPLICANT: Bergeron, John J.M.
; APPLICANT: Thomas, David Y.
; TITLE OF INVENTION: Method for screening for
; TITLE OF INVENTION: UDP-glucose:glycoprotein glucosyltransferase (UGGT) activity
; TITLE OF INVENTION: and nucleic acid encoding for UGGT
; FILE REFERENCE: 2139-9"US"
; CURRENT APPLICATION NUMBER: US/09/376,330
; CURRENT FILING DATE: 1999-08-18
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 26
; LENGTH: 107
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: GLUT-S. typ
```

US-09-376-330-26

Query Match 2.0%; Score 6; DB 4; Length 107;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 KDVDSI 77  
DB 37 KDVDSI 42

RESULT 20

US-07-634-278-47  
; Sequence 47, Application US/07634278  
; Patent No. 5530101  
; GENERAL INFORMATION:  
; APPLICANT: QUEEN, Cary L.  
; APPLICANT: CO, Man Sung  
; APPLICANT: SCHNEIDER, William P.  
; APPLICANT: LANDOLFI, Nicholas F.  
; APPLICANT: COELINGH, Kathleen L.  
; APPLICANT: SELICK, Harold E.  
; TITLE OF INVENTION: IMPROVED HUMANIZED IMMUNOGLOBULINS  
; NUMBER OF SEQUENCES: 113  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Khourie and Crew  
; STREET: 379 Lytton Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: US  
; ZIP: 94301

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/634,278  
FILING DATE: 19-DEC-1990  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/590,274  
FILING DATE: 28-SEP-1990

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/310,252  
FILING DATE: 13-FEB-1989

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/290,975  
FILING DATE: 28-DEC-1988

ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William M  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 11823-002600

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 326-2422

INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 111 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

US-07-634-278-47

Query Match 2.0%; Score 6; DB 1; Length 111;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 GTRVEI 124  
DB 105 GTRVEI 110

RESULT 21

US-08-477-728-47  
; Sequence 47, Application US/08477728  
; Patent No. 5585089  
; GENERAL INFORMATION:  
; APPLICANT: QUEEN, Cary L.  
; APPLICANT: SCHNEIDER, William P.  
; APPLICANT: SELICK, Harold E.  
; TITLE OF INVENTION: IMPROVED HUMANIZED IMMUNOGLOBULINS  
; NUMBER OF SEQUENCES: 113  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: US  
; ZIP: 94111

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/477,728  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/634,278  
FILING DATE: 19-DEC-1990

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/590,274  
FILING DATE: 28-SEP-1990

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/310,252  
FILING DATE: 13-FEB-1989

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/290,975  
FILING DATE: 28-DEC-1988

ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William M  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 11823-002600

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 326-2422

INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 111 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

US-08-477-728-47

Query Match 2.0%; Score 6; DB 1; Length 111;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 GTRVEI 124  
DB 105 GTRVEI 110

RESULT 22

US-08-474-040-47  
; Sequence 47, Application US/08474040  
; Patent No. 5693761  
; GENERAL INFORMATION:  
; APPLICANT: QUEEN, Cary L.  
; APPLICANT: CO, Man Sung  
; APPLICANT: SCHNEIDER, William P.  
; APPLICANT: LANDOLFI, Nicholas F.



```

; APPLICANT: COELINGH, Kathleen L.
; APPLICANT: SELICK, Harold E.
; TITLE OF INVENTION: IMPROVED HUMANIZED IMMUNOGLOBULINS
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: 379 Lytton Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: US
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,040
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/634,278
; FILING DATE: 19-DEC-1990
; APPLICATION NUMBER: US 07/590,274
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/310,252
; FILING DATE: 13-FEB-1989
; APPLICATION NUMBER: US 07/290,975
; FILING DATE: 28-DEC-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, William M
; REGISTRATION NUMBER: 30,223
; REFERENCE/DOCKET NUMBER: 11823-002600
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 326-2400
; TELEFAX: (415) 326-2422
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 111 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-474-040-47

Query Match          2.0%; Score 6; DB 1; Length 111;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      119 GTRVEI 124
Db      105 GTRVEI 110

RESULT 23
US-08-487-200-47
; Sequence 47, Application US/08487200
; Patent No. 5693762
; GENERAL INFORMATION:
; APPLICANT: QUEEN, Cary L.
; APPLICANT: CO, Man Sung
; APPLICANT: SCHNEIDER, William P.
; APPLICANT: LANDOLFI, Nicholas F.
; APPLICANT: COELINGH, Kathleen L.
; APPLICANT: SELICK, Harold E.
; TITLE OF INVENTION: IMPROVED HUMANIZED IMMUNOGLOBULINS
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: 379 Lytton Avenue
; CITY: Palo Alto

```

```

; STATE: California
; COUNTRY: US
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,200
; FILING DATE: 7-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/634,278
; FILING DATE: 19-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/590,274
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/310,252
; FILING DATE: 13-FEB-1989
; APPLICATION NUMBER: US 07/290,975
; FILING DATE: 28-DEC-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, William M
; REGISTRATION NUMBER: 30,223
; REFERENCE/DOCKET NUMBER: 11823-002610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 326-2400
; TELEFAX: (415) 326-2422
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 111 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-487-200-47

Query Match          2.0%; Score 6; DB 1; Length 111;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      119 GTRVEI 124
Db      105 GTRVEI 110

RESULT 24
US-08-484-537-47
; Sequence 47, Application US/08484537
; Patent No. 6180370
; GENERAL INFORMATION:
; APPLICANT: QUEEN, Cary L.
; APPLICANT: CO, Man Sung
; APPLICANT: SCHNEIDER, William P.
; APPLICANT: LANDOLFI, Nicholas F.
; APPLICANT: COELINGH, Kathleen L.
; APPLICANT: SELICK, Harold E.
; TITLE OF INVENTION: IMPROVED HUMANIZED IMMUNOGLOBULINS
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: 379 Lytton Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: US
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/484,537  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/634,278  
;; FILING DATE: 19-DEC-1990  
;; APPLICATION NUMBER: US/07/590,274  
;; FILING DATE: 28-SEP-1990  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/310,252  
;; FILING DATE: 13-FEB-1989  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/290,975  
;; FILING DATE: 28-DEC-1988  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Smith, William M  
;; REGISTRATION NUMBER: 30,223  
;; REFERENCE/DOCKET NUMBER: 11823-002600  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 326-2400  
;; TELEFAX: (415) 326-2422  
;; INFORMATION FOR SEQ ID NO: 47:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 111 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
US-08-484-537-47

Query Match 2.0%; Score 6; DB 4; Length 111;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 GTRVEI 124  
Db 105 GTRVEI 110

RESULT 25  
US-08-983-607-25  
; Sequence 25, Application US/08983607  
; Patent No. 6140470  
; GENERAL INFORMATION:  
; APPLICANT: Alan Garen  
; APPLICANT: Xiaohong Cai  
; TITLE OF INVENTION: Human Anti-Tumor Monoclonal Anti-  
; TITLE OF INVENTION: bodies  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Department of Molecular Biophysics  
; ADDRESSEE: and Biochemistry, Yale University  
; STREET: 266 Whitney Avenue  
; CITY: New Haven  
; STATE: Connecticut  
; COUNTRY: United States of America  
; ZIP: 06520-8114  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" 1.44 Mb diskette  
; COMPUTER: IBM PC  
; OPERATING SYSTEM: MS DOS  
; SOFTWARE: Word Processing  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/983,607  
; FILING DATE: April 27, 1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/IB96/01032  
; FILING DATE: June 28, 1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

;; NAME: Mary M. Krinsky  
;; REGISTRATION NUMBER: 32423  
;; REFERENCE/DOCKET NUMBER: OCR-679  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 203-773-9544  
;; TELEFAX: 203-773-1183  
;; INFORMATION FOR SEQ ID NO: 25:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 113 residues  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: polypeptide  
;; DESCRIPTION: polypeptide  
;; ORIGINAL SOURCE:  
;; ORGANISM: Homo sapiens (melanoma patient immu-  
;; ORGANISM: nized with autologous tumor cells)  
;; INDIVIDUAL ISOLATE: peripheral blood lymphocytes  
;; IMMEDIATE SOURCE:  
;; LIBRARY: DM414 scFv antibodies obtained from  
;; CLONE: D33  
;; FEATURE:  
;; NAME/KEY: light chain  
;; OTHER INFORMATION: Xaa at position 47 is His or  
;; OTHER INFORMATION: Gly and Xaa at position 95 is His or Gln  
US-08-983-607-25

Query Match 2.0%; Score 6; DB 4; Length 113;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 GTRVEI 124  
Db 106 GTRVEI 111

RESULT 26  
US-09-045-764A-6  
; Sequence 6, Application US/09045764A  
; Patent No. 6127178  
; GENERAL INFORMATION:  
; APPLICANT: Israel, Mark A.  
; APPLICANT: Florio, Monica  
; TITLE OF INVENTION: Apoptotic Peptides  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 75 DENISE DRIVE  
; CITY: HILLSBOROUGH  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94010  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/045,764A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, RICHARD A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: UCSF98-045  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 343-4341  
; TELEFAX: (650) 343-4342  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 128 amino acids  
; TYPE: amino acid

STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-09-045-764A-6

Query Match 2.0%; Score 6; DB 3; Length 128;  
 Best Local Similarity 100.0%; Pred. No. 3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 263 AELASI 268  
 Db 109 AELASI 114

## RESULT 27

US-08-936-165A-481  
 ; Sequence 481, Application US/08936165A  
 ; Patent No. 6348582

## GENERAL INFORMATION:

APPLICANT: Black, Michael  
 APPLICANT: Burnham, Martin  
 APPLICANT: Hodgson, John  
 APPLICANT: Knowles, David  
 APPLICANT: Lonetto, Michael  
 APPLICANT: Nicholas, Richard  
 APPLICANT: Pratt, Julie  
 APPLICANT: Reichard, Richard  
 APPLICANT: Rosenberg, Martin  
 APPLICANT: Ward, Judith

TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,  
 TITLE OF INVENTION: Polypeptides and Their Uses

NUMBER OF SEQUENCES: 534

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation  
 STREET: 709 Swedeland Road  
 CITY: King of Prussia  
 STATE: PA

COUNTRY: USA

ZIP: 19406-0939

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/936.165A

FILING DATE: 24-SEP-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/027,032

FILING DATE: 24-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Gimmi, Edward R

REGISTRATION NUMBER: 38,891

REFERENCE/DOCKET NUMBER: P50549

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-4478

TELEFAX: 610-270-5090

TELEX:

INFORMATION FOR SEQ ID NO: 481:

SEQUENCE CHARACTERISTICS:

LENGTH: 148 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Protein.

US-08-936-165A-481

Query Match 2.0%; Score 6; DB 4; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 ISSKA 28

Db 95 ISSKA 100

## RESULT 28

US-08-936-165A-389  
 ; Sequence 389, Application US/08936165A  
 ; Patent No. 6348582

## GENERAL INFORMATION:

APPLICANT: Black, Michael  
 APPLICANT: Burnham, Martin  
 APPLICANT: Hodgson, John  
 APPLICANT: Knowles, David  
 APPLICANT: Lonetto, Michael  
 APPLICANT: Nicholas, Richard  
 APPLICANT: Pratt, Julie  
 APPLICANT: Reichard, Richard  
 APPLICANT: Rosenberg, Martin  
 APPLICANT: Ward, Judith

TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,  
 TITLE OF INVENTION: Polypeptides and Their Uses

NUMBER OF SEQUENCES: 534

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation  
 STREET: 709 Swedeland Road  
 CITY: King of Prussia  
 STATE: PA

COUNTRY: USA

ZIP: 19406-0939

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/936.165A

FILING DATE: 24-SEP-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/027,032

FILING DATE: 24-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Gimmi, Edward R

REGISTRATION NUMBER: 38,891

REFERENCE/DOCKET NUMBER: P50549

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-4478

TELEFAX: 610-270-5090

TELEX:

INFORMATION FOR SEQ ID NO: 389:

SEQUENCE CHARACTERISTICS:

LENGTH: 156 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Protein

US-08-936-165A-389

Query Match 2.0%; Score 6; DB 4; Length 156;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 LLSLPN 22

Db 75 LLSLPN 80

US-08-936-165A-389

## RESULT 29

US-09-010-809-18  
 ; Sequence 18, Application US/09010809B  
 ; Patent No. 6090801

## GENERAL INFORMATION:

APPLICANT: Gustafsson, Claes

```

; APPLICANT: Betlach, Mary C.
; TITLE OF INVENTION: Epithelone Polyketide Synthases and Encoding DNA
; FILE OF INVENTION: Therefor
; FILE REFERENCE: 30062-20020.00
; CURRENT APPLICATION NUMBER: US/09/010,809B
; CURRENT FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-010-809-18

Query Match          2.0%; Score 6; DB 3; Length 158;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 273 SAVATL 278
DB 64 SAVATL 69
|||||

RESULT 30
US-08-597-236-9
; Sequence 9, Application US/08597236
; Patent No. 5733765
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesca
; APPLICANT: MOLLET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; TITLE OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/597,236
; FILING DATE: 20-JUN-1995
; CLASSIFICATION: 426
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95201669.9
; FILING DATE: 20-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci A., Allan
; REGISTRATION NUMBER: 30256
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 159 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-597-236-9

Query Match          2.0%; Score 6; DB 1; Length 159;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 GARLTF 293
DB 30 GARLTF 35
|||||

RESULT 31
US-08-597-236-9
; Sequence 9, Application US/08746682A
; Patent No. 5786184
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesca
; APPLICANT: MOLLET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; TITLE OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,682A
; FILING DATE: 14-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/597,236
; FILING DATE: 20-JUN-1995
; APPLICATION NUMBER: EP 95201669.9
; FILING DATE: 20-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci A., Allan
; REGISTRATION NUMBER: 30256
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 159 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-746-682A-9

Query Match          2.0%; Score 6; DB 1; Length 159;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 GARLTF 293
DB 30 GARLTF 35
|||||

RESULT 32
US-09-615-192A-318
; Sequence 318, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
; APPLICANT: Blosberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; TITLE OF INVENTION: Modification of Plant Lignin Content
; FILE REFERENCE: 11000.1003c4U
; CURRENT APPLICATION NUMBER: US/09/615,192A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 405
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 318
; LENGTH: 161
; TYPE: PRT
```

```
; ORGANISM: Eucalyptus grandis
US-09-615-192A-318

Query Match      2.0%; Score 6; DB 4; Length 161;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      148 FALARE 153
Db      72 FALARE 77

RESULT 33
US-09-370-838-41
; Sequence 41, Application US/09370838
; Patent No. 644425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-370-838-41

Query Match      2.0%; Score 6; DB 4; Length 163;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      83 ASVGIS 88
Db      87 ASVGIS 92

RESULT 34
US-08-858-207A-353
; Sequence 353, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328el Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 353:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 167 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6348328e
US-08-858-207A-353

Query Match      2.0%; Score 6; DB 4; Length 167;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      80 KTDASV 85
Db      77 KTDASV 82

RESULT 35
US-09-149-476-354
; Sequence 354, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
```

[illegible]

```

; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/057,650
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/056,884
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,669
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/049,610
; EARLIER FILING DATE: 1997-06-13
; EARLIER APPLICATION NUMBER: 60/061,060
; EARLIER FILING DATE: 1997-10-02

Query Match          2.0%; Score 6; DB 4; Length 169;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 AVFQDN 104
    |||||
Db 88 AVFQDN 93

RESULT 36
US-09-134-001C-3134
; Sequence 3134, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3134
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3134

Query Match          2.0%; Score 6; DB 4; Length 170;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 29 INNAK 34
    |||||
Db 127 INNAK 132

RESULT 37
US-09-134-001C-3854
; Sequence 3854, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3854
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3854

Query Match          2.0%; Score 6; DB 4; Length 173;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 179 LSIISV 184
    |||||
Db 122 LSIISV 127

RESULT 38
US-09-495-066-2
; Sequence 2, Application US/09495066
; Patent No. 6358708
; GENERAL INFORMATION:
; APPLICANT: Scriptgen Pharmaceuticals, Inc.
; APPLICANT: Craig M. Thompson
; TITLE OF INVENTION: CANDIDA ALBICANS SRB-7
; FILE REFERENCE: 0342/1F124-US1
; CURRENT APPLICATION NUMBER: US/09/495,066
; CURRENT FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/118, 604
; PRIOR FILING DATE: 1999-02-04
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-495-066-2

Query Match          2.0%; Score 6; DB 4; Length 178;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 161 PKEKVS 166
    |||||
Db 89 PKEKVS 94

RESULT 39
US-09-134-001C-4050
; Sequence 4050, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4050
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4050

Query Match          2.0%; Score 6; DB 4; Length 198;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 242 ASLYYH 247
    |||||
Db 48 ASLYYH 53

RESULT 40
```

```

US-09-497-779A-6
; Sequence 6, Application US/09497779A
; Patent No. 6297368
; GENERAL INFORMATION:
; APPLICANT: DEPINHO, RONALD A.
; TITLE OF INVENTION: MYC HOMOLOG REGION II - ASSOCIATED PROTEIN AND USES THEREOF
; FILE REFERENCE: 96700-609
; CURRENT APPLICATION NUMBER: US/09/497,779A
; CURRENT FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 08/946,692
; PRIOR FILING DATE: 1997-10-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 199
; TYPE: PRT
; ORGANISM: MOUSE
US-09-497-779A-6

```

```

Query Match      2.0%; Score 6; DB 4; Length 199;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      74 VDSIET 79
      |||||
Db      94 VDSIET 99

```

```

Search completed: July 8, 2003, 10:34:17
Job time : 48 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 10:20:35 ; Search time 72 Seconds  
(without alignments)  
542.256 Million cell updates/sec

Title: US-10-062-624-40

Perfect score: 293

Sequence: 1 MNKLKTTINTVLVCLLSL.....AVATLNTGYFGGIGARLTF 293

Scoring table:

OLIGO Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : A\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	293	100.0	293	20	AAU06965
2	293	100.0	293	23	AAU96115
3	25	8.5	280	20	AAU06949
4	25	8.5	293	23	AAU73412
5	24	8.2	292	20	AAU06967
6	22	7.5	298	23	AAU73410
7	17	5.8	271	20	AAU06958
8	17	5.8	300	23	AAU73411
9	16	5.5	164	20	AAU06951
10	16	5.5	291	23	AAU73409
					E. canis P30-5 pro
					Ehrlichia canis p2
					E. chafeensis omp
					Ehrlichia chafeen
					E. canis P30-7 pro
					Ehrlichia chafeen
					E. chafeensis omp
					Ehrlichia chafeen
					E. chafeensis omp
					Ehrlichia chafeen

11	15	5.1	275	20	AAU06956	E. chafeensis omp
12	15	5.1	275	23	AAU73407	Ehrlichia chafeen
13	13	4.4	161	20	AAU06969	E. canis P30-9 pro
14	11	3.8	279	20	AAU06954	E. chafeensis omp
15	11	3.8	279	23	AAU73405	Ehrlichia chafeen
16	11	3.8	283	20	AAU06955	E. chafeensis omp
17	11	3.8	283	23	AAU73406	Ehrlichia chafeen
18	11	3.8	295	20	AAU06953	E. chafeensis omp
19	11	3.8	295	23	AAU73404	Ehrlichia chafeen
20	11	3.8	299	20	AAU06968	E. canis P30-8 pro
21	10	3.4	285	20	AAU06957	E. chafeensis omp
22	10	3.4	285	23	AAU73408	Ehrlichia chafeen
23	10	3.4	289	20	AAU06986	E. canis P30-6 pro
24	8	2.7	131	20	AAU06952	E. chafeensis omp
25	8	2.7	272	23	AAU73403	Ehrlichia chafeen
26	8	2.7	280	23	AAU96116	Ehrlichia canis p2
27	7	2.4	65	23	ABP10881	Human ORFX protein
28	7	2.4	74	22	AAU53609	Protonibacterium
29	7	2.4	101	22	ABBA4342	Peptide #10848 enc
30	7	2.4	101	22	AAU64250	Human brain expres
31	7	2.4	101	22	AAU77074	Human bone marrow
32	7	2.4	101	22	AAU37196	Peptide #11233 enc
33	7	2.4	115	21	AAU21513	Arabidopsis thalia
34	7	2.4	128	22	ABU70338	Drosophila melanog
35	7	2.4	158	18	AAW22483	Phaffia derived gl
36	7	2.4	182	20	AAU60163	Human endometrium
37	7	2.4	209	22	ABG24630	Novel human diagno
38	7	2.4	249	22	ABG29708	Novel human diagno
39	7	2.4	253	21	AAU58276	Arabidopsis thalia
40	7	2.4	253	21	AAU61124	Arabidopsis thalia
41	7	2.4	274	23	ABP40434	Staphylococcus epi
42	7	2.4	279	22	AAU29942	Human protein sequ
43	7	2.4	281	23	ABP41568	Human ovarian anti
44	7	2.4	283	20	AAU06963	E. canis P30-3 pro
45	7	2.4	283	21	AAU71478	Ehrlichia canis im
46	7	2.4	283	22	ABG25902	Novel human diagno
47	7	2.4	283	23	AAU96101	Ehrlichia canis p2
48	7	2.4	292	22	AAU98968	E. coli growth and
49	7	2.4	303	21	AAU45702	Arabidopsis thalia
50	7	2.4	315	21	AAU95719	Cosmid CHRIM5 enco
51	7	2.4	326	23	ABG60923	Novel floral maris
52	7	2.4	327	21	AAU45701	Arabidopsis thalia
53	7	2.4	334	23	ABU54639	Lactococcus lactis
54	7	2.4	336	13	AAU28149	Sugar beet beta-1,
55	7	2.4	343	22	ABU70069	Drosophila melanog
56	7	2.4	399	23	ABP41360	Human ovarian anti
57	7	2.4	411	22	ABG14795	Novel human diagno
58	7	2.4	424	22	ABG44620	Drosophila melanog
59	7	2.4	467	21	AAU45700	Arabidopsis thalia
60	7	2.4	467	23	ABU93586	Herbicidally activ
61	7	2.4	472	22	AAU79011	C. glutamicum SRT
62	7	2.4	486	22	ABG14794	Novel human diagno
63	7	2.4	505	22	AAU33352	Human keratin enco
64	7	2.4	505	22	AAU33354	Human breast cance
65	7	2.4	518	22	AAU91327	C glutamicum prote
66	7	2.4	523	23	ABU78800	Human NOV2b protei
67	7	2.4	532	22	ABU12372	Human bone marrow
68	7	2.4	551	21	AAU52398	Human keratin KERT
69	7	2.4	551	23	AAU20423	Human keratin-2 (K
70	7	2.4	581	22	ABU12455	Human bone marrow
71	7	2.4	629	22	AAU02721	C-terminally trunc
72	7	2.4	663	23	ABU93753	Herbicidally activ
73	7	2.4	900	20	AAU30338	A multifunctional
74	7	2.4	900	22	AAU20176	Saccharomyces cere
75	7	2.4	900	22	AAU20177	S. cerevisiae mult
76	7	2.4	900	22	AAU20178	S. cerevisiae mult
77	7	2.4	900	22	AAU20179	S. cerevisiae mult
78	7	2.4	982	22	ABU71714	Drosophila melanog
79	7	2.4	1013	22	ABU58267	Drosophila melanog
80	7	2.4	1025	21	AAU15926	E. coli proliferat
81	7	2.4	1025	22	AAU98973	E. coli growth and
82	7	2.4	1046	23	ABU54658	Lactococcus lactis
83	7	2.4	1098	22	ABG30367	Novel human diagno

84	7	2.4	1487	21	ARB23826	Murine phospholipase	157	6	2.0	61	23	ABP05669	Human ORFX protein
85	7	2.4	1501	22	AG82878	S. epidermidis ope	158	6	2.0	62	21	AGG10285	Arabidopsis thalia
86	7	2.4	1529	23	ABG319100	Staphylococcus epi	159	6	2.0	62	21	ABY64772	Human 5' EST relat
87	7	2.4	1677	22	ABG110016	Novel human diagno	160	6	2.0	62	22	ABB29479	Peptide #2130 enco
88	7	2.4	1720	22	AAW78886	Human protein SEQ	161	6	2.0	62	22	ABB34658	Peptide #2164 enco
89	7	2.4	1721	22	AAW79870	Human protein SQ	162	6	2.0	62	22	ABB20064	Protein #2063 enco
90	7	2.4	2090	22	ABB64682	Drosophila melanog	163	6	2.0	62	22	AAW55447	Human brain expres
91	7	2.4	3587	14	AAK34713	Bacillus subtilis	164	6	2.0	62	22	AAW55454	Peptide #2088 enco
92	6	2.0	10	19	AAW77116	Tyrosinase synthet	165	6	2.0	62	22	AAW28155	Peptide #2192 enco
93	6	2.0	10	22	AAW42903	Mycoplasma genital	166	6	2.0	62	22	AAW03391	Peptide #2073 enco
94	6	2.0	10	22	AAW42937	Mycoplasma genital	167	6	2.0	62	22	ABP03086	Human ORFX protein
95	6	2.0	10	22	AAW17644	Human PAPalpha spe	168	6	2.0	62	23	ABP07062	Human ORFX protein
96	6	2.0	13	7	AP60993	Sequence of immuno	169	6	2.0	62	23	ABP07062	Human m3 muscarini
97	6	2.0	16	23	AAE14476	Oplophorus gracili	170	6	2.0	68	17	AAW03583	Novel human secret
98	6	2.0	20	13	AAE21805	AntiAChR antibody	171	6	2.0	68	22	AAU33146	Human ORF3354 prot
99	6	2.0	24	23	AAE14484	Oplophorus lucifer	172	6	2.0	69	22	ABP34381	Human polypeptide
100	6	2.0	30	13	AAU27734	HindB4R. Vaccinia	173	6	2.0	69	22	AAO09430	Human polypeptide
101	6	2.0	30	23	AAU84903	Human Tyros segmen	174	6	2.0	69	23	ABP27676	Streptococcus poly
102	6	2.0	30	23	AAU84904	Human Tyros segmen	175	6	2.0	70	23	ABP202304	Human ORFX protein
103	6	2.0	30	23	AAU84990	Human MUC1R segmen	176	6	2.0	72	22	AAO10076	Human polypeptide
104	6	2.0	30	23	AAU84991	Human MUC1R segmen	177	6	2.0	73	21	AAU35583	Arabidopsis thalia
105	6	2.0	35	21	AAAB45245	Human secreted pro	178	6	2.0	73	22	AAU58982	Propionibacterium
106	6	2.0	38	22	ABE28755	Peptide #1406 enco	179	6	2.0	77	22	AAU64906	Propionibacterium
107	6	2.0	38	22	ABE33942	Peptide #1448 enco	180	6	2.0	80	16	AAW70741	S. typhimurium prg
108	6	2.0	38	22	ABE19379	Protein #1378 enco	181	6	2.0	80	18	AAW18386	Lung cancer associ
109	6	2.0	38	22	AAW54705	Human brain expres	182	6	2.0	81	21	AAW58164	Homologous sequenc
110	6	2.0	38	22	AAW67105	Human bone marrow	183	6	2.0	82	17	AAU33000	Arabidopsis thalia
111	6	2.0	38	22	AAW14967	Peptide #1401 enco	184	6	2.0	83	21	AAW24280	Group B Streptococ
112	6	2.0	38	22	AAW27401	Peptide #1438 enco	185	6	2.0	83	21	AAW91305	Arabidopsis thalia
113	6	2.0	38	22	AAW2692	Peptide #1374 enco	186	6	2.0	83	22	AAU21001	Human novel foetal
114	6	2.0	38	23	ABG36765	Human peptide enco	187	6	2.0	84	23	ABP05957	Human ORFX protein
115	6	2.0	41	20	AAW06777	Amino acid fragmen	188	6	2.0	86	22	AAW87536	Human immune/haema
116	6	2.0	46	22	AAW87773	Human immune/haema	189	6	2.0	88	23	ABP22206	Human ORF1179 prot
117	6	2.0	46	23	ABE90720	Chlamydia pneumoni	190	6	2.0	88	23	ABP27028	Streptococcus poly
118	6	2.0	47	22	ABE03678	Human musculoskele	191	6	2.0	91	22	ABG28763	Propionibacterium
119	6	2.0	48	22	ABE30382	Peptide #3033 enco	192	6	2.0	91	22	ABP40582	Novel human diagno
120	6	2.0	48	22	ABE35555	Peptide #3061 enco	193	6	2.0	93	22	AAU57434	Staphylococcus epi
121	6	2.0	48	22	ABE20980	Protein #2979 enco	194	6	2.0	96	22	ABG21811	Propionibacterium
122	6	2.0	48	22	AAW87744	Human bone marrow	195	6	2.0	96	22	AAU32639	Novel human diagno
123	6	2.0	48	22	AAW16570	Peptide #3004 enco	196	6	2.0	96	22	AAU32639	Novel human secret
124	6	2.0	48	22	AAW29056	Peptide #3093 enco	197	6	2.0	98	23	ABP08187	Human ORFX protein
125	6	2.0	48	22	AAW04285	Peptide #2967 enco	198	6	2.0	99	22	ABE330590	Peptide #3241 enco
126	6	2.0	48	23	ABG38329	Human peptide enco	199	6	2.0	99	22	AAW68946	Human bone marrow
127	6	2.0	49	22	ABE13260	Human nervous syst	200	6	2.0	99	22	AAW16770	Peptide #3204 enco
128	6	2.0	50	23	ABP02311	Human ORFX protein	201	6	2.0	99	22	AAW29257	Peptide #3294 enco
129	6	2.0	52	21	AAW05444	Arabidopsis thalia	202	6	2.0	99	22	AAW04487	Peptide #3169 enco
130	6	2.0	52	21	AAW53693	Arabidopsis thalia	203	6	2.0	100	21	ABG38530	Human peptide enco
131	6	2.0	52	22	AAU20845	Human novel foetal	204	6	2.0	100	22	AAW64671	Human 5' EST relat
132	6	2.0	53	22	ABE10807	Human ovarian and/	205	6	2.0	100	22	AAW68627	Human immune/haema
133	6	2.0	53	22	AAW94473	Human reproductive	206	6	2.0	103	22	AAU54041	Propionibacterium
134	6	2.0	53	22	AAW85962	Human immune/haema	207	6	2.0	103	22	AAU03721	Group B Streptococ
135	6	2.0	54	22	ABE96347	Human testicular a	208	6	2.0	103	22	AAW63436	Human breast cance
136	6	2.0	54	22	AAW95816	Human reproductive	209	6	2.0	103	23	ABP09540	Human ORFX protein
137	6	2.0	55	22	ABE28623	Peptide #1274 enco	210	6	2.0	104	22	AAU62491	Propionibacterium
138	6	2.0	55	22	ABE13806	Peptide #1312 enco	211	6	2.0	105	22	AAU65851	Propionibacterium
139	6	2.0	55	22	ABE17013	Human nervous syst	212	6	2.0	105	22	ABP076756	Human colon cancer
140	6	2.0	55	22	ABE19249	Protein #1248 enco	213	6	2.0	105	23	ABP07937	Human ORFX protein
141	6	2.0	55	22	AAW54572	Human brain expres	214	6	2.0	106	22	AAW87474	Novel central nerv
142	6	2.0	55	22	AAW66979	Human bone marrow	215	6	2.0	107	22	AAW72454	Partial GLUT. Uni
143	6	2.0	55	22	AAW14839	Peptide #1273 enco	216	6	2.0	108	21	AAW52192	Human anti-HBs ant
144	6	2.0	55	22	AAW27266	Peptide #1303 enco	217	6	2.0	108	23	ABP34099	Human kinase-like
145	6	2.0	55	22	AAW02564	Peptide #1246 enco	218	6	2.0	108	23	AAO16958	Human myoglobulin
146	6	2.0	55	23	ABG36636	Human peptide enco	219	6	2.0	109	23	ABP03286	Human ORFX protein
147	6	2.0	55	23	ABP43039	Human ovarian anti	220	6	2.0	111	13	AAW5721	Humanised VL regio
148	6	2.0	58	21	AAW51964	Human secreted pro	221	6	2.0	111	22	AAW69662	Humanised Fd/9 ant
149	6	2.0	58	21	AAW34329	Human secreted pro	222	6	2.0	112	23	AAU12167	Human partial lyso
150	6	2.0	58	22	AAW74819	Human colon cancer	223	6	2.0	113	18	AAW13517	Anti-melanoma ligh
151	6	2.0	58	23	ABP11094	Human ORFX protein	224	6	2.0	114	22	AAU32572	Novel human secret
152	6	2.0	60	20	AAW60403	Human normal bladd	225	6	2.0	115	20	AAW35427	Chlamydia pneumoni
153	6	2.0	60	22	ABG14201	Novel human diagno	226	6	2.0	115	22	AAW80214	Human autoantibody
154	6	2.0	61	21	AAW00399	Human secreted pro	227	6	2.0	116	23	AAW48215	Listeria monocytoc
155	6	2.0	61	21	AAW01072	Human secreted pro	228	6	2.0	117	23	AAW43334	Human ORFX ORF3098
156	6	2.0	61	21	AAW01667	Human secreted pro	229	6	2.0	118	20	AAW27337	Group B Streptococ
										118	22	AAW84775	Human immune/haema

230	6	2.0	118	23	RAO14050	Human anti-FAPalpa	303	6	2.0	165	21	AAG30703	Arabidopsis thalia
231	6	2.0	118	23	ABP30133	Streptococcus poly	304	6	2.0	165	22	ABG12135	Novel human diagno
232	6	2.0	120	22	AAU67063	Propionibacterium	305	6	2.0	166	21	AGG08846	Arabidopsis thalia
233	6	2.0	120	23	ABP42597	Human ovarian anti	306	6	2.0	166	21	AAG52222	Arabidopsis thalia
234	6	2.0	121	21	AAG01891	Human secreted pro	307	6	2.0	166	23	ABP09066	Human ORFX protein
235	6	2.0	121	22	ABG95469	Human secreted pro	308	6	2.0	167	19	AAW38557	Streptococcus pneu
236	6	2.0	122	21	AAG01808	Human protein sequ	309	6	2.0	167	22	AAG75451	Human colon cancer
237	6	2.0	122	22	AAU32234	Human secreted pro	310	6	2.0	168	21	AAW95745	Bacillus megaterium
238	6	2.0	124	21	AAG11657	Novel human secret	311	6	2.0	169	19	AAW74765	Human secreted pro
239	6	2.0	124	23	ABP02824	Arabidopsis thalia	312	6	2.0	170	22	ABG13566	Novel human diagno
240	6	2.0	126	22	AAU52151	Propionibacterium	313	6	2.0	170	23	ABP38289	Staphylococcus epi
241	6	2.0	126	23	ABW7064	AC2885 antibody li	314	6	2.0	171	22	ABP38289	Human gene 21 enco
242	6	2.0	127	22	RAO01535	Human polypeptide	315	6	2.0	171	23	ABW7434	Streptococcus pneu
243	6	2.0	127	23	ABP29772	Streptococcus poly	316	6	2.0	173	18	AAW11244	Streptococcus pneu
244	6	2.0	128	21	AAW19321	Amino acid sequenc	317	6	2.0	173	22	ABW7403	Human gene 21 enco
245	6	2.0	128	21	AAW58485	Arabidopsis thalia	318	6	2.0	173	23	ABG65391	Human albumin fusi
246	6	2.0	128	22	AAG71997	Human olfactory re	319	6	2.0	173	23	ABP39009	Staphylococcus epi
247	6	2.0	130	22	AAO07175	Human polypeptide	320	6	2.0	176	10	AAW90414	Plasmidium falcipa
248	6	2.0	130	23	ABP031007	Human ORFX protein	321	6	2.0	176	21	ABG25095	Arabidopsis thalia
249	6	2.0	132	21	AAG12886	Arabidopsis thalia	322	6	2.0	176	22	ABG1759	S. epidermidis ope
250	6	2.0	133	22	AAU34779	E. coli cellular p	323	6	2.0	178	21	AAW08339	Amino acid sequenc
251	6	2.0	133	22	AAU38214	Salmonella typhi c	324	6	2.0	180	13	AAW27664	C-terminal region
252	6	2.0	133	22	AAG98988	E. coli growth and	325	6	2.0	180	23	ABP27937	Streptococcus poly
253	6	2.0	135	21	AAG11656	Arabidopsis thalia	326	6	2.0	182	21	ABW48337	Listeria monocytog
254	6	2.0	135	22	AAU41859	Propionibacterium	327	6	2.0	182	21	ABW58484	Arabidopsis thalia
255	6	2.0	135	23	ABP33053	Human nuclease-lik	328	6	2.0	183	15	AAW50974	Human calcivirius S
256	6	2.0	136	22	AAU60138	Propionibacterium	329	6	2.0	183	23	ABP02634	Human ORFX ORF655
257	6	2.0	137	22	ABG01697	Novel human diagno	330	6	2.0	185	22	AAW25959	Human protein sequ
258	6	2.0	137	22	ABG24199	Novel human diagno	331	6	2.0	185	22	ABG03230	Human ovarian and/
259	6	2.0	137	22	AAW94082	Human reproductive	332	6	2.0	186	22	ABG03230	Human diagno
260	6	2.0	137	23	ABW50061	Listeria monocytog	333	6	2.0	187	22	ABW10926	Human reproductiv
261	6	2.0	138	22	AAU42948	Propionibacterium	334	6	2.0	187	22	AAW96362	Human prostate can
262	6	2.0	139	23	ABP32038	Human glycoprotein	335	6	2.0	189	21	ABW57233	Escherichia coli p
263	6	2.0	140	22	AAU32607	Novel human enzyme	336	6	2.0	189	22	ABW52524	Human polypeptide
264	6	2.0	140	22	AAU25586	Human G Protein-Co	337	6	2.0	189	23	ABW90090	Herbicidally activ
265	6	2.0	140	22	AAO05855	Human polypeptide	338	6	2.0	191	21	ABW91121	Novel central nerv
266	6	2.0	141	21	AAG33700	Arabidopsis thalia	339	6	2.0	192	22	AAW87290	Arabidopsis thalia
267	6	2.0	141	22	AAU87159	Novel central nerv	340	6	2.0	193	21	AAG08845	Arabidopsis thalia
268	6	2.0	142	20	AAW73987	Human prostate tum	341	6	2.0	193	21	AAG52221	Arabidopsis thalia
269	6	2.0	144	21	AAW35473	Pinus radiata cell	342	6	2.0	195	21	AAG33698	Arabidopsis thalia
270	6	2.0	145	22	AAG82617	S. epidermidis ope	343	6	2.0	196	21	ABW56837	Human prostate can
271	6	2.0	146	20	AAW41718	Human PRO862 prote	344	6	2.0	197	21	AAG23855	Arabidopsis thalia
272	6	2.0	146	21	AAW44274	Human PRO862 (UNQ4	345	6	2.0	197	21	AAG53776	Arabidopsis thalia
273	6	2.0	146	21	AAG08847	Arabidopsis thalia	346	6	2.0	198	23	ABP39205	Staphylococcus epi
274	6	2.0	146	21	AAG52223	Arabidopsis thalia	347	6	2.0	199	21	AAG08426	Arabidopsis thalia
275	6	2.0	146	21	AAW57398	Human lysoenzyme L	348	6	2.0	199	23	AAU10385	Murine MHR11-AP62
276	6	2.0	146	22	AAU67959	Propionibacterium	349	6	2.0	200	19	AAW98629	H. pylori GHPO 286
277	6	2.0	146	22	AAU12375	Human PRO862 polyp	350	6	2.0	204	20	AAW36806	Protein which is s
278	6	2.0	146	22	AAG81278	Human AFP protein	351	6	2.0	207	19	AAW38549	S. pneumoniae pept
279	6	2.0	147	21	AAG25203	Arabidopsis thalia	352	6	2.0	211	20	AAW35201	. Protein involved i
280	6	2.0	147	21	AAG61718	Arabidopsis thalia	353	6	2.0	211	22	AAW29467	Human.G protein-co
281	6	2.0	147	22	ABW11115	Novel human diagno	354	6	2.0	211	23	ABG60755	Novel G protein co
282	6	2.0	148	19	AAW77721	Formate dehydrogen	355	6	2.0	215	21	AAW32576	Human secreted pro
283	6	2.0	148	21	ABW33112	Pinus radiata tran	356	6	2.0	215	22	ABW44707	Amino acid sequenc
284	6	2.0	150	21	ABW33179	Eucalyptus grandis	357	6	2.0	215	22	ABW44709	Amino acid sequenc
285	6	2.0	151	22	AAW87362	Human gene 21 enco	358	6	2.0	215	22	AAW75362	Human secreted pro
286	6	2.0	151	23	ABG65390	Human albumin fusi	359	6	2.0	216	21	AAW71022	Human Mucin 1 (MUC
287	6	2.0	152	23	ABP34021	Human ORF2994 prot	360	6	2.0	216	22	AAG72067	Human olfactory re
288	6	2.0	154	21	AAW27961	Human secreted pro	361	6	2.0	217	22	ABG09524	Novel human diagno
289	6	2.0	154	21	AAW37714	The kappa chain of	362	6	2.0	217	22	AAU01529	Corn Wuschel (WUS)
290	6	2.0	155	22	ABW64163	Drosophila melanog	363	6	2.0	217	22	AAW87629	Bovine mammary tis
291	6	2.0	155	22	AAW77629	Primosomal protein	364	6	2.0	218	22	AAU54440	Propionibacterium
292	6	2.0	156	23	AAU80560	Human G protein-co	365	6	2.0	221	21	AAW63215	Gene 39 human secr
293	6	2.0	157	22	AAU64993	Propionibacterium	366	6	2.0	222	20	AAW02661	Human secreted pro
294	6	2.0	159	18	AAW14077	S.thermophilus exo	367	6	2.0	222	22	AAW90710	C Glutamicum prote
295	6	2.0	159	18	AAW22176	S.thermophilus exo	368	6	2.0	223	22	AAU23010	Novel human enzyme
296	6	2.0	159	21	AAW1655	Arabidopsis thalia	369	6	2.0	229	22	AAU30974	Novel human secret
297	6	2.0	161	21	AAW16392	Eucalyptus grandis	370	6	2.0	236	23	ABW43147	Human ovarian anti
298	6	2.0	162	22	ABG15991	Novel human diagno	371	6	2.0	237	23	ABP29051	Streptococcus poly
299	6	2.0	163	20	AAW29526	Human lung tumour	372	6	2.0	240	22	AAW69737	Human Muts homolog
300	6	2.0	163	21	AAW44416	Human lung tumour-	373	6	2.0	240	22	ABW79417	Corynebacterium g1
301	6	2.0	163	21	AAW33699	Arabidopsis thalia	374	6	2.0	240	23	ABW27775	Streptococcus poly
302	6	2.0	163	22	AAW13757	Human lung tumour-	375	6	2.0	245	23	ABW49160	Listeria monocytog

376	6	2.0	247	22	ABB63083	Drosophila melanog	449	295	23	ABP30459	Streptococcus poly
377	6	2.0	248	23	ABP22776	Streptococcus poly	450	295	23	ABBS0266	Human polypeptide
378	6	2.0	249	21	AAW29244	Arabidopsis thalia	451	295	23	ABBS0113	Listeria monocytog
379	6	2.0	250	18	AAW25146	Luffin-B (a riboso	452	299	22	AAU36982	Staphylococcus aur
380	6	2.0	250	18	AAW21709	Luffin-B. luffa c	453	299	22	ABP41229	Human ovarian anti
381	6	2.0	250	21	AAW18126	Arabidopsis thalia	454	300	19	AAW76187	Actinoplanes sp. a
382	6	2.0	251	22	ABBS9026	Drosophila melanog	455	300	23	AAW49554	Actinoplanes sp ac
383	6	2.0	251	22	ABBS6776	Drosophila melanog	456	301	21	AAW53246	Human colon cancer
384	6	2.0	253	23	AAO14057	Anti-FAPalpha anti	457	301	22	AAW74071	Human colon cancer
385	6	2.0	254	22	AAW90927	C glutamic prote	458	302	23	AAU95771	Human olfactory an
386	6	2.0	254	23	ABP45458	Human BLYS binding	459	302	22	ABG19711	Novel human diago
387	6	2.0	255	22	ABG17132	Novel human diago	460	306	21	AAW42217	Human ORFX ORF1981
388	6	2.0	255	23	AAO14054	Anti-PAPalpha anti	461	307	21	AAW08331	Arabidopsis thalia
389	6	2.0	256	23	ABBS91271	Herbicidally activ	462	307	21	AAW51930	Arabidopsis thalia
390	6	2.0	257	22	AAW09435	Arabidopsis thalia	463	308	22	AAW26276	S. epidermidis ope
391	6	2.0	257	22	AAU56436	Protonibacterium	464	308	23	ABP28206	Streptococcus poly
392	6	2.0	258	21	AAW93223	An Escherichia col	465	308	23	ABP28206	Nicotiana tabacum
393	6	2.0	258	23	ABP40748	Staphylococcus epi	466	309	22	AAW01017	CPE 16 protein seq
394	6	2.0	258	23	AAU10387	Murine MHR11-AP62	467	309	22	AAW82638	S. epidermidis ope
395	6	2.0	262	23	AAO14058	Anti-FAPalpha anti	468	310	22	AAU34420	E. coli cellular p
396	6	2.0	263	21	AAW05592	Arabidopsis thalia	469	310	23	ABU04043	Human G-protein co
397	6	2.0	263	21	AAW51377	C. jejuni flgg pro	470	311	21	AAW06945	Arabidopsis thalia
398	6	2.0	263	22	AAW81696	S. epidermidis ope	471	312	22	ABG09525	Novel human diago
399	6	2.0	263	23	ABP38217	Staphylococcus epi	472	312	23	ABP38613	Streptococcus epi
400	6	2.0	263	23	ABP29315	Streptococcus poly	473	312	23	ABP30334	Streptococcus poly
401	6	2.0	265	22	AAU38258	Pseudomonas aerugi	474	312	23	ABP30334	Human MUC1R consen
402	6	2.0	265	22	AAW62363	Human methionyl tr	475	312	21	AAU94810	Arabidopsis thalia
403	6	2.0	266	23	AAW98518	Micrococcus lysae	476	313	22	AAW20288	Arabidopsis thalia
404	6	2.0	267	22	ABW76806	2,5-diketo-D-gluc	477	313	22	AAW99845	AGC protein kinase
405	6	2.0	267	22	AAE04387	Human P2-purinerigi	478	313	22	AAW82332	S. epidermidis ope
406	6	2.0	268	23	ABW07890	Homo DNA topoisome	479	315	12	AAW13159	Herbicidally activ
407	6	2.0	269	10	AAW91444	Region of typtoph	480	315	22	AAE04385	2-ketocaldonate red
408	6	2.0	269	23	ABP29417	Streptococcus poly	481	315	22	AAW72503	Human P2-purinerigi
409	6	2.0	271	18	AAW27704	B. pertussis BrKa	482	316	22	ABG13568	Human OR-like poly
410	6	2.0	271	22	AAE00445	HCV E2-634 HVR1 co	483	317	22	AAW72230	Novel human diago
411	6	2.0	271	22	AAE00446	HCV E2-634 HVR1-mu	484	317	22	AAW72382	Human olfactory re
412	6	2.0	271	23	AAW96118	Ehrlichia canis p2	485	317	22	ABP28704	Human OR-like poly
413	6	2.0	273	15	AAW48368	Human tyrosinase.	486	317	23	ABP28704	Streptococcus poly
414	6	2.0	273	22	AAU39169	Propionibacterium	487	318	21	AAW92507	Human polypeptide
415	6	2.0	274	22	ABW70933	Drosophila melanog	488	318	22	AAU14224	Human OXRE-4 with
416	6	2.0	275	23	ABJ04008	Human G-protein co	489	318	22	AAU14460	Human novel protei
417	6	2.0	275	22	AAU28046	Novel human secret	490	318	22	AAW95401	Human protein sequ
418	6	2.0	277	19	AAW98512	H. pylori GHPO 127	491	318	22	AAW64556	Human secreted pro
419	6	2.0	277	22	ASG02134	Novel human diago	492	318	22	AAW64604	Human secreted pro
420	6	2.0	277	23	ABW90568	Chlamydia pneumoni	493	318	23	AAW14254	Human 23155 protei
421	6	2.0	278	13	AAW29910	Prod. of the luffi	494	319	18	AAW27715	S. marcesens Sep a
422	6	2.0	278	22	AAW82909	S. epidermidis ope	495	319	18	AAW27716	S. marcesens Sep a
423	6	2.0	279	23	ABG28048	Novel human diago	496	319	21	AAW92913	Arabidopsis thalia
424	6	2.0	279	23	ABW50065	Listeria monocytog	497	319	23	ABW92913	Herbicidally activ
425	6	2.0	280	14	AAW43055	Insecticidal/Fungi	498	320	18	AAW27717	S. marcesens Sep-h
426	6	2.0	280	23	ABW9765	Herbicidally activ	499	320	18	AAW27718	S. marcesens Sep-h
427	6	2.0	282	21	AAW23732	Arabidopsis thalia	500	321	21	AAW43416	Human cancer assoc
428	6	2.0	282	21	AAW51907	Arabidopsis thalia	501	321	21	ABG18805	Novel human diago
429	6	2.0	283	20	AAW06944	E. chafeensis OMP-	502	325	21	AAW06659	Arabidopsis thalia
430	6	2.0	283	23	AAU96106	Ehrlichia chafeens	503	325	22	AAW72506	Alternative versio
431	6	2.0	283	23	AAU73413	Ehrlichia chafeen	504	326	21	AAW08299	Arabidopsis thalia
432	6	2.0	284	21	AAW14375	Arabidopsis thalia	505	327	15	AAW96298	Glycoprotein 39 C
433	6	2.0	284	21	AAW04166	Arabidopsis thalia	506	327	21	AAW29684	Arabidopsis thalia
434	6	2.0	285	21	AAW05591	Arabidopsis thalia	507	327	22	AAW71444	Human olfactory re
435	6	2.0	285	23	AAU73402	Ehrlichia chafeen	508	327	23	ABW49967	Listeria monocytog
436	6	2.0	287	13	AAW27665	Secreted form of H	509	328	23	ABP38384	Staphylococcus epi
437	6	2.0	287	21	AAW40516	Arabidopsis thalia	510	328	23	ABW90218	Human polypeptide
438	6	2.0	287	22	AAW46413	H. pylori HPC180 p	511	329	20	AAW92181	Human RCE1 (hRCE1)
439	6	2.0	290	22	ABG18137	Novel human diago	512	330	21	AAW5966	Rat skin cell prot
440	6	2.0	291	23	ABP40050	Staphylococcus epi	513	330	21	AAW76023	Rat skin cell prot
441	6	2.0	292	22	ABW70355	Drosophila melanog	514	330	22	AAW64603	Human secreted pro
442	6	2.0	292	22	ABG19274	Novel human diago	515	330	22	AAW55905	Skin cell protein,
443	6	2.0	292	23	ABW97372	Novel human protei	516	330	22	AAW55962	Skin cell protein,
444	6	2.0	294	22	AAU33765	Staphylococcus aur	517	330	23	ABW72105	Rat protein isolat
445	6	2.0	295	20	AAW37169	Protein involved i	518	330	23	ABW72162	Rat protein isolat
446	6	2.0	295	21	AAW23731	Arabidopsis thalia	519	331	21	AAW96257	R. eutropha toluen
447	6	2.0	295	21	AAW51906	Arabidopsis thalia	520	331	23	ABW93941	Herbicidally activ
448	6	2.0	295	22	ABW52495	Escherichia coli p	521	331	23	ABW84727	DNA polymerase III

522	6	2.0	333	21	AAB25682	Human secreted pro	595	6	2.0	356	23	ABB47847	Listeria monocytoc
523	6	2.0	333	22	AAU41730	Propionibacterium	596	6	2.0	357	22	ABG61145	Drosophila melanog
524	6	2.0	333	22	ABG28216	Novel human diagn	597	6	2.0	357	22	ABG61145	Novel human diagn
525	6	2.0	337	21	AAU70082	S. cerevisiae vana	598	6	2.0	358	20	AAU28280	Chlamydia pneumoni
526	6	2.0	337	21	AAG92054	C glutamicum prote	599	6	2.0	358	20	AAU35652	Chlamydia pneumoni
527	6	2.0	337	22	AAB79386	Corynebacterium gl	600	6	2.0	358	21	AAU05958	Protein deduced fr
528	6	2.0	338	21	AAG08330	Arabisopsis thalia	601	6	2.0	358	21	AAU45039	Human G-protein-co
529	6	2.0	338	21	AAG31323	Arabisopsis thalia	602	6	2.0	358	21	AAU44365	Human G-protein co
530	6	2.0	338	21	AAG51929	Arabisopsis thalia	603	6	2.0	358	22	AAE03199	Human G-protein co
531	6	2.0	338	21	AAU71028	Arabisopsis thalia	604	6	2.0	358	22	AAE02499	Human G-protein co
532	6	2.0	338	21	AAU59426	Ubiquitin-E. coli	605	6	2.0	358	22	AAE02499	Human G-protein co
533	6	2.0	338	22	ABG24202	Soybean inositol 1	606	6	2.0	358	22	AAU96884	Human novel G-prot
534	6	2.0	338	22	ABG24202	Novel human diagn	607	6	2.0	358	22	AAU96884	Human novel G-prot
535	6	2.0	339	18	AAU55578	H. pylori ORF 06ep	608	6	2.0	358	22	AAU96884	Human G-protein co
536	6	2.0	339	20	AAU86309	Kidney injury asso	609	6	2.0	358	22	AAU96884	Human G-protein co
537	6	2.0	339	21	AAU58194	Human STRAP-1 prot	610	6	2.0	358	22	AAU96884	Human G-protein co
538	6	2.0	339	22	ABG26163	Novel human diagn	611	6	2.0	358	22	AAU96884	Human G-protein co
539	6	2.0	339	22	AAU69927	Human prostate cDN	612	6	2.0	359	23	AAU31322	Arabisopsis thalia
540	6	2.0	339	22	AAU78845	Human protein SEQ	613	6	2.0	359	23	AAU31322	Arabisopsis thalia
541	6	2.0	339	22	AAU01282	P789P amino acid s	614	6	2.0	360	16	AAU79166	Oplophorus gracilli
542	6	2.0	339	23	ABG61813	Prostate cancer-as	615	6	2.0	360	16	AAU79166	Oplophorus gracilli
543	6	2.0	339	23	ABG5387	Human P789P protei	616	6	2.0	360	18	AAU35833	Human monocyte che
544	6	2.0	340	19	AAU98578	H. pylori GHPO 49	617	6	2.0	360	22	ABG56340	Non-endogenous hum
545	6	2.0	341	20	AAU35341	Chlamydia pneumoni	618	6	2.0	360	22	AAU80108	Human CCR2b protei
546	6	2.0	341	22	AAU24529	Human olfactory re	619	6	2.0	360	22	AAU28047	Novel human secret
547	6	2.0	341	22	AAU81566	S. epidermidis ope	620	6	2.0	360	22	AAU07613	Human CCR2-641 pol
548	6	2.0	341	22	AAU72499	Human SNORF40 rece	621	6	2.0	360	22	AAU07613	Human wild-type CC
549	6	2.0	342	21	AAU81576	G-coupled olfacto	622	6	2.0	361	21	AAU33774	Arabisopsis thalia
550	6	2.0	342	21	AAU71306	EBV-induced G-prot	623	6	2.0	361	21	AAU33774	Arabisopsis thalia
551	6	2.0	342	21	AAU40368	Human orphan G pro	624	6	2.0	361	22	AAU76588	C glutamicum prote
552	6	2.0	342	21	AAU02840	Arabisopsis thalia	625	6	2.0	362	21	AAU06658	Arabisopsis thalia
553	6	2.0	342	21	AAU94444	Human G protein co	626	6	2.0	362	21	AAU06658	Arabisopsis thalia
554	6	2.0	342	21	AAU94444	Human 15625 recept	627	6	2.0	363	19	AAU75788	Staphylococcus aur
555	6	2.0	342	22	AAU96643	Macaque ortholog o	628	6	2.0	363	21	AAU47271	Arabisopsis thalia
556	6	2.0	342	22	AAU96643	Putative P. abyss	629	6	2.0	363	22	AAU37208	Arabisopsis thalia
557	6	2.0	342	22	AAU79249	Human protein SEQ	630	6	2.0	364	21	AAU06944	Staphylococcus aur
558	6	2.0	342	22	AAE04386	Human P2-purinergi	631	6	2.0	365	21	AAU06657	Arabisopsis thalia
559	6	2.0	342	23	ABU77868	Amino acid sequenc	632	6	2.0	365	23	AAO20961	Arabisopsis thalia
560	6	2.0	342	23	AAU80164	Human ADP-glucose	633	6	2.0	367	21	AAU16523	Arabisopsis thalia
561	6	2.0	342	23	AAU48353	Human G-protein-co	634	6	2.0	367	21	AAU47617	Arabisopsis thalia
562	6	2.0	342	23	ABU05031	Human SPI68 recept	635	6	2.0	368	21	AAU06943	Arabisopsis thalia
563	6	2.0	342	23	ABU05031	Human PAFR3 protei	636	6	2.0	369	22	ABG24784	Novel human diagn
564	6	2.0	343	21	AAG08329	Arabisopsis thalia	637	6	2.0	371	22	ABG58533	Drosophila melanog
565	6	2.0	343	21	AAG23730	Arabisopsis thalia	638	6	2.0	371	22	ABG66446	B thuringiensis 44
566	6	2.0	343	21	AAG51905	Arabisopsis thalia	639	6	2.0	372	21	AAU87481	Arabisopsis thalia
567	6	2.0	343	21	AAG51928	Arabisopsis thalia	640	6	2.0	372	21	AAU35818	Arabisopsis thalia
568	6	2.0	343	22	AAE04384	Rat P2-purinergic	641	6	2.0	373	21	AAU26450	Arabisopsis thalia
569	6	2.0	344	21	AAU40367	Arabisopsis thalia	642	6	2.0	373	22	AAU31662	S. epidermidis ope
570	6	2.0	346	21	AAU23854	Arabisopsis thalia	643	6	2.0	374	16	AAU79165	Human monocyte che
571	6	2.0	346	21	AAU53775	Arabisopsis thalia	644	6	2.0	374	21	AAU35817	Arabisopsis thalia
572	6	2.0	346	22	ABG06006	Novel human diagn	645	6	2.0	374	21	AAU47269	Arabisopsis thalia
573	6	2.0	348	13	AAU27662	Human protein kina	646	6	2.0	374	22	AAU80107	Human CCR2a protei
574	6	2.0	348	22	ABU31165	C-terminal region	647	6	2.0	374	22	AAU79829	Human protein SEQ
575	6	2.0	348	22	ABU31165	Peptide #3816 enco	648	6	2.0	374	22	AAU99820	AGC protein kinase
576	6	2.0	348	22	ABU31165	Peptide #3866 enco	649	6	2.0	375	21	AAU06906	Arabisopsis thalia
577	6	2.0	348	22	ABU21716	Protein #3715 enco	650	6	2.0	375	21	AAU26700	Arabisopsis thalia
578	6	2.0	348	22	AAU57127	Human brain expres	651	6	2.0	375	21	AAU42848	Arabisopsis thalia
579	6	2.0	348	22	AAU69519	Human bone marrow	652	6	2.0	375	21	AAU47293	Arabisopsis thalia
580	6	2.0	348	22	AAU29858	Peptide #3895 enco	653	6	2.0	375	21	AAU47293	Arabisopsis thalia
581	6	2.0	348	23	ABG31444	Human peptide enco	654	6	2.0	377	22	AAU02780	Human six transmem
582	6	2.0	349	23	ABG92879	Herbicidally activ	655	6	2.0	377	22	AAU27595	Fusion protein 730
583	6	2.0	349	21	AAU35483	Arabisopsis thalia	656	6	2.0	377	22	AAU75543	Human B7-like prot
584	6	2.0	350	21	AAU47614	Arabisopsis thalia	657	6	2.0	379	23	ABU54011	Lactococcus lactis
585	6	2.0	350	21	AAU24667	Arabisopsis thalia	658	6	2.0	380	23	ABU54505	Lactococcus lactis
586	6	2.0	350	22	ABU50338	Arabisopsis thalia	659	6	2.0	382	22	AAU70002	Rat G-protein coup
587	6	2.0	350	22	ABU6843	Drosophila melanog	660	6	2.0	384	13	AAU27508	Human RBP (GRP-R)
588	6	2.0	350	22	ABU05170	Novel human diagn	661	6	2.0	384	22	ABU56359	Non-endogenous hum
589	6	2.0	350	22	AAE05986	Clostridium diffi	662	6	2.0	384	22	AAU84205	Amino acid sequenc
590	6	2.0	351	11	AAU06024	Viral membrane pro	663	6	2.0	385	18	AAU33708	B. thuringiensis 4
591	6	2.0	352	21	AAU08298	Arabisopsis thalia	664	6	2.0	385	20	AAU35524	Chlamydia pneumoni
592	6	2.0	353	22	AAU27596	Fusion protein 730	665	6	2.0	385	21	AAU19136	Amino acid sequenc
593	6	2.0	354	19	AAU54037	Mouse CC-CKS prot	666	6	2.0	385	22	AAU87462	B thuringiensis 45
594	6	2.0	355	21	AAU40515	Arabisopsis thalia	667	6	2.0	385	22	AAU87483	B thuringiensis 44
595	6	2.0	356	21	AAU08297	Arabisopsis thalia							

668	6	2.0	386	21	AA226699	Arabidopsis thalia	741	6	2.0	424	23	AAE21706	Human PKIN-1 prote
669	6	2.0	386	21	AA226699	Arabidopsis thalia	742	6	2.0	424	23	AAU10023	Human protein kina
670	6	2.0	386	23	AB55321	Lactococcus lactis	743	6	2.0	425	22	ABU62832	Drosophila melanog
671	6	2.0	387	21	AA94498	Rat MP-10 receptor	744	6	2.0	426	21	AAAG3356	Arabidopsis thalia
672	6	2.0	388	23	ABP38975	Staphylococcus epi	745	6	2.0	426	21	AAAG4280	Arabidopsis thalia
673	6	2.0	388	23	ABP28890	Streptococcus poly	746	6	2.0	426	22	ABB61821	Drosophila melanog
674	6	2.0	389	22	AA33840	Staphylococcus aur	747	6	2.0	426	23	ABB47703	Listeria monocytog
675	6	2.0	391	22	AA894155	Human protein sequ	748	6	2.0	426	23	ABP38040	Staphylococcus epi
676	6	2.0	391	22	AA887488	B thuringiensis 44	749	6	2.0	429	22	AAU12218	Human PRO4331 poly
677	6	2.0	392	22	AA887494	B thuringiensis 44	750	6	2.0	431	14	AA833931	Morphogen hof1 ful
678	6	2.0	393	21	AA814374	Arabidopsis thalia	751	6	2.0	433	21	AA858755	Breat and ovarian
679	6	2.0	393	21	AA816522	Arabidopsis thalia	752	6	2.0	433	22	AA82838	C glutamicum prote
680	6	2.0	393	21	AA816522	Arabidopsis thalia	753	6	2.0	433	23	ABP77560	Corynebacteria Cys
681	6	2.0	393	21	AA816522	Arabidopsis thalia	754	6	2.0	435	23	ABP30041	Streptococcus poly
682	6	2.0	393	21	AA816522	Arabidopsis thalia	755	6	2.0	436	23	ABP30041	Human octamer bind
683	6	2.0	393	21	AA816522	Arabidopsis thalia	756	6	2.0	436	23	ABP30041	Human octamer bind
684	6	2.0	395	21	AA804117	Human protein sequ	757	6	2.0	440	20	AA817179	H. pylori outer me
685	6	2.0	395	21	AA804117	Human protein sequ	758	6	2.0	442	22	ABP68120	Novel human protei
686	6	2.0	395	22	AA876714	phyltyl/prenyltrans	759	6	2.0	443	18	AAW21014	Drosophila melanog
687	6	2.0	395	22	AA865799	Corynebacterium gl	760	6	2.0	443	20	AAW40497	H. pylori cytoplas
688	6	2.0	395	22	AA865799	RPP5-like protein	761	6	2.0	443	20	AAW40497	H. influenzae vael
689	6	2.0	395	23	AAU72785	Soybean ATP2 hom	762	6	2.0	443	20	AAW40497	S. aureus priA pro
690	6	2.0	396	12	AA813720	Human mevalonate k	763	6	2.0	443	21	AAW40497	Arabidopsis thalia
691	6	2.0	398	22	AAU36619	Staphylococcus aur	764	6	2.0	444	22	ABP71087	Drosophila melanog
692	6	2.0	398	22	AAU37319	Staphylococcus aur	765	6	2.0	444	22	ABP71087	Drosophila melanog
693	6	2.0	398	22	AAU37469	Staphylococcus aur	766	6	2.0	444	22	ABP71087	Drosophila melanog
694	6	2.0	398	22	AA884475	S. epidermidis ope	767	6	2.0	446	23	ABP28722	Herbicidally activ
695	6	2.0	398	22	AA884475	S. epidermidis ope	768	6	2.0	446	23	ABP28722	Herbicidally activ
696	6	2.0	399	13	AA827511	Human R3BP. Homo	769	6	2.0	450	21	AAW90584	Streptococcus poly
697	6	2.0	399	21	AAW98919	Human bombesin rec	770	6	2.0	450	21	AAW90584	Streptococcus poly
698	6	2.0	399	22	AB56333	Non-endogenous hum	771	6	2.0	452	21	AAW90584	Arabidopsis thalia
699	6	2.0	399	22	AB47512	Bombesin receptor	772	6	2.0	452	21	AAW90584	Arabidopsis thalia
700	6	2.0	399	22	AAU07638	Rat bombesin-like	773	6	2.0	452	21	AAW90584	Arabidopsis thalia
701	6	2.0	400	18	AAW22981	Canine herpesvirus	774	6	2.0	453	20	AAW97049	Protein encoded by
702	6	2.0	400	19	AAW22648	Canine herpesvirus	775	6	2.0	453	20	AAW97049	Sphingomonas sp. s
703	6	2.0	400	21	AAW90209	Arabidopsis thalia	776	6	2.0	453	22	AAW92444	Sphingomonas desul
704	6	2.0	400	21	AAW90209	Arabidopsis thalia	777	6	2.0	453	22	AAW92444	Sphingomonas desul
705	6	2.0	400	22	ABG13840	Novel human diagno	778	6	2.0	454	22	AAU35138	Human secreted pro
706	6	2.0	400	22	AB513305	Canine herpes viru	779	6	2.0	455	12	AAU35138	Enterococcus faeca
707	6	2.0	400	23	ABP92276	Herbicidally activ	780	6	2.0	455	13	AAU35138	Human 55KD TNF-bin
708	6	2.0	401	21	AAW28161	Herbicidally activ	781	6	2.0	455	22	AAU35138	Transmembrane form
709	6	2.0	402	23	ABP91758	Herbicidally activ	782	6	2.0	455	22	AAU35138	Human Mucin 1 (MUC
710	6	2.0	402	23	ABP91758	Herbicidally activ	783	6	2.0	455	22	AAU35138	S. epidermidis ope
711	6	2.0	403	21	AB08640	Herbicidally activ	784	6	2.0	456	21	AAU35138	S. epidermidis ope
712	6	2.0	403	22	AAW90150	C glutamicum prote	785	6	2.0	456	21	AAU35138	Arabidopsis thalia
713	6	2.0	404	23	AAW90150	Arabidopsis thalia	786	6	2.0	456	21	AAU35138	Arabidopsis thalia
714	6	2.0	404	23	AAW90150	Arabidopsis thalia	787	6	2.0	457	21	AAU35138	Drosophila melanog
715	6	2.0	404	23	ABP91757	Herbicidally activ	788	6	2.0	458	22	AAU35138	Arabidopsis thalia
716	6	2.0	404	23	ABP91757	Herbicidally activ	789	6	2.0	459	21	AAU35138	Arabidopsis thalia
717	6	2.0	404	23	ABP92059	Herbicidally activ	790	6	2.0	459	22	AAU35138	Arabidopsis thalia
718	6	2.0	407	21	AAU35138	Arabidopsis thalia	791	6	2.0	460	22	AAU35138	Arabidopsis thalia
719	6	2.0	407	21	AAU35138	Arabidopsis thalia	792	6	2.0	461	23	ABP38471	Drosophila melanog
720	6	2.0	408	21	AAW90605	Arabidopsis thalia	793	6	2.0	462	23	ABP38471	Arabidopsis thalia
721	6	2.0	408	21	AAW90605	Arabidopsis thalia	794	6	2.0	462	23	ABP38471	Arabidopsis thalia
722	6	2.0	408	21	AAW93768	The HtrA surface p	795	6	2.0	463	22	ABP38471	Arabidopsis thalia
723	6	2.0	411	23	AAU72786	Soybean ATP2 hom	796	6	2.0	463	22	ABP38471	Arabidopsis thalia
724	6	2.0	415	22	AAU72786	Propionibacterium	797	6	2.0	465	22	AAU34285	Staphylococcus aur
725	6	2.0	417	21	AAW33904	E. coli cellular p	798	6	2.0	465	22	AAU34285	S. epidermidis ope
726	6	2.0	417	22	AAU34572	Human polypeptide,	799	6	2.0	466	23	ABP38471	Herbicidally activ
727	6	2.0	417	22	AAW93526	Human polypeptide,	800	6	2.0	467	21	AAU75499	Neisseria meningit
728	6	2.0	417	23	ABP38965	Staphylococcus epi	801	6	2.0	467	21	AAU75499	Neisseria meningit
729	6	2.0	418	22	AAU3594	Propionibacterium	802	6	2.0	468	22	AAU3597	Staphylococcus aur
730	6	2.0	418	22	AAU3594	Propionibacterium	803	6	2.0	468	22	AAU3597	S. epidermidis ope
731	6	2.0	419	22	AAU56910	Propionibacterium	804	6	2.0	468	23	ABP38471	Herbicidally activ
732	6	2.0	419	22	ABG20438	Novel human diagno	805	6	2.0	469	18	AAW23820	Human sarcoclectin
733	6	2.0	419	22	AAW6713	Corynebacterium gl	806	6	2.0	469	21	AAW6713	Amino acid sequenc
734	6	2.0	421	21	AAW42846	Arabidopsis thalia	807	6	2.0	470	22	AAU52137	Protonibacterium
735	6	2.0	421	22	AAW39514	Propionibacterium	808	6	2.0	470	19	AAW98287	H. pylori GHPO 144
736	6	2.0	422	21	AAW69504	Arabidopsis thalia	809	6	2.0	471	19	AAW98287	H. pylori GHPO 701
737	6	2.0	423	20	AAW52238	Amino acid sequenc	810	6	2.0	472	22	ABG20228	Protein involved i
738	6	2.0	424	22	AAW52238	Mouse skp2 SEQ ID	811	6	2.0	472	22	ABG20228	Novel human diagno
739	6	2.0	424	22	AAU03544	Human protein kina	812	6	2.0	473	23	AAE09508	Human mucin-1 (MUC
740	6	2.0	424	23	AAU79585	Novel human protei	813	6	2.0	473	23	ABP38719	Staphylococcus epi

814	6	2.0	474	23	ABP38396	Staphylococcus epi	887	19	AAW59448	Pea ADP-glucose py
815	6	2.0	475	21	AAW50583	Arabidopsis thalia	888	21	AAG20281	Arabidopsis thalia
816	6	2.0	476	22	AAW50573	Human MUC1 polypep	889	21	AAG49884	Arabidopsis thalia
817	6	2.0	477	22	ABW77476	Human MUC1 SEQ ID	890	21	ABG31986	Arabidopsis thalia
818	6	2.0	478	22	AAW60749	Gene 2 related pep	891	21	ABG31684	Herbicidally activ
819	6	2.0	479	22	AAW60750	Gene 2 related pep	892	21	ABG40402	Novel human diagno
820	6	2.0	480	22	AAW37305	Amino acid sequenc	893	22	ABG04082	Novel human diagno
821	6	2.0	481	19	AAW60239	Rose 1-aminocyclop	894	22	AAR92483	Human protein sequ
822	6	2.0	482	23	ABW53703	Lactococcus lactis	895	22	AAW36717	Laccase-LCCI, Pol
823	6	2.0	483	21	AAW97730	Alpha-L-arabinofur	896	20	AAW36717	Arabidopsis thalia
824	6	2.0	483	21	AAW12130	Hydrophobic domain	897	20	ABW77510	Trametes versicolo
825	6	2.0	483	21	AAW08639	A murine TANGO 201	898	21	ABW78802	Human NOV2d protei
826	6	2.0	483	21	AAW08646	A murine TANGO 201	899	21	AAW07238	Human polypeptide
827	6	2.0	483	21	AAW08647	A murine TANGO 201	900	22	AAW41006	Staphylococcus aur
828	6	2.0	483	21	AAW08648	A murine TANGO 201	901	22	AAW34406	C glutamicum prote
829	6	2.0	483	21	AAW08649	A murine TANGO 201	902	22	AAG91296	Human DITHP polype
830	6	2.0	483	21	AAW08650	A murine TANGO 201	903	23	ABG60120	Human DITHP polype
831	6	2.0	483	21	AAW08651	A murine TANGO 201	904	22	AAW19582	Arabidopsis thalia
832	6	2.0	483	21	AAW08652	A murine TANGO 201	905	22	AAW19582	Human diagnostic a
833	6	2.0	483	21	AAW08653	A murine TANGO 201	906	22	AAW56309	Human tyrosinase a
834	6	2.0	483	21	AAW08654	A murine TANGO 201	907	22	AAR63623	Human tyrosinase p
835	6	2.0	483	21	AAW08655	A murine TANGO 201	908	22	AAW03306	Tyrosinase melanom
836	6	2.0	483	21	AAW95005	Human secreted pro	909	22	AAW00184	Human tyrosinase
837	6	2.0	483	22	AAW28141	Novel human secret	910	22	AAW36519	Tyrosinase contain
838	6	2.0	483	22	AAW62406	Human MSP10 polyp	911	18	AAW38166	Normal human tyros
839	6	2.0	483	22	AAW93456	Human protein sequ	912	19	AAW71234	Tyrosinase, Synth
840	6	2.0	483	22	AAW60471	Human cell cycle a	913	22	AAW86039	Human tyrosinase p
841	6	2.0	483	23	AAW10794	Human zinc finger	914	22	AAW51344	Human tyrosinase p
842	6	2.0	484	21	AAW53734	Arabidopsis thalia	915	23	ABW83482	Human cytoskeleton
843	6	2.0	488	23	ABW64400	Drosophila melanog	916	23	AAW4806	Human Tyros consen
844	6	2.0	488	23	ABW48920	Listeria monocytog	917	23	AAW11544	Human Melanoma tum
845	6	2.0	489	14	AAW44295	Corticosterone-bin	918	21	AAW05942	Protein deduced fr
846	6	2.0	489	21	AAW97820	Pseudomonas sp. WF	919	22	AAG40002	Arabidopsis thalia
847	6	2.0	490	22	AAW37569	Staphylococcus aur	920	22	AAG78732	Hamster oculoctan
848	6	2.0	493	22	AAW46224	Propionibacterium	921	16	AAW07071	Sequence oocoduct b
849	6	2.0	495	22	AAW00539	Human MUC1 polypep	922	18	AAW79493	Human tyrosinase a
850	6	2.0	497	22	ABW68854	Novel human diagno	923	18	AAW22083	Human SK29-MEL tyr
851	6	2.0	497	22	AAW00443	HCV HVRI-7 constru	924	22	AAW53526	Propionibacterium
852	6	2.0	497	22	AAW00444	HCV HVRI-7 mut5 cons	925	22	AAG81726	S. epidermidis ope
853	6	2.0	498	21	AAW08480	Amino acid sequenc	926	20	AAW23338	Aspergillus niger
854	6	2.0	498	21	AAW75500	Neisseria meningit	927	20	AAW18090	Truncated A. niger
855	6	2.0	499	11	AAW05279	Amino acid sequenc	928	20	AAW03450	Aspergillus niger
856	6	2.0	499	11	AAW07069	Phenol oxidase (PO	929	21	AAW77740	A. niger G2 glucosa
857	6	2.0	499	11	AAW07068	Phenol oxidase (PO	930	22	AAW61904	A. niger G2 glucosa
858	6	2.0	499	19	AAW76295	Polyporus pinsitus	931	22	AAW48171	A. niger G1 glucosa
859	6	2.0	499	19	AAW76296	Polyporus pinsitus	932	22	ABW64613	Drosophila melanog
860	6	2.0	499	19	AAW76297	Polyporus pinsitus	933	22	AAW33628	Yeast transcrip
861	6	2.0	499	19	AAW76298	Polyporus pinsitus	934	23	ABW26404	Streptococcus poly
862	6	2.0	499	19	AAW76299	Polyporus pinsitus	935	23	ABW60984	Novel human protei
863	6	2.0	499	21	AAW60875	Polyporus pinsitus	936	23	AAW98586	Blatrid beetle lu
864	6	2.0	499	21	AAW50582	Arabidopsis thalia	937	22	AAU34682	E. coli cellular p
865	6	2.0	501	13	AAW53733	Arabidopsis thalia	938	21	AAW52397	Human keratin KERT
866	6	2.0	501	13	AAW27741	Sequence transcrib	939	23	AAW20422	Human keratin-1 (K
867	6	2.0	501	22	AAW98428	Escherichia coli p	940	23	ABW64798	Human albumin fusi
868	6	2.0	502	21	AAG4542	Arabidopsis thalia	941	23	AAW96185	Human secreted pro
869	6	2.0	502	21	AAG32392	Arabidopsis thalia	942	22	AAG49530	Arabidopsis thalia
870	6	2.0	504	21	AAW20282	Arabidopsis thalia	943	22	ABG25521	Novel human diagno
871	6	2.0	504	21	AAW49885	Arabidopsis thalia	944	22	ABW50168	Human transcrip
872	6	2.0	504	23	ABW41965	Human ovarian anti	945	18	AAW27150	HMG-CoA reductase
873	6	2.0	505	21	AAW31987	Arabidopsis thalia	946	21	AAG53315	Arabidopsis thalia
874	6	2.0	505	22	ABW71008	Drosophila melanog	947	23	ABW66695	Human novel polype
875	6	2.0	507	14	AAW42284	Protein which lowe	948	23	ABW92296	Herbicidally activ
876	6	2.0	508	18	AAW38167	Mutant human tyros	949	23	ABW91439	Herbicidally activ
877	6	2.0	508	19	AAW77233	MiniMUC1 protein.	950	23	ABW60985	Novel human protei
878	6	2.0	509	21	AAW10702	Arabidopsis thalia	951	21	ABW29941	Streptococcus poly
879	6	2.0	511	22	AAW38272	Salmonella typhi c	952	21	AAG31546	Arabidopsis thalia
880	6	2.0	511	22	AAW92220	Human polypeptide	953	21	AAG16507	Arabidopsis thalia
881	6	2.0	512	22	AAW60703	Human secreted pro	954	9	AAW31156	Human tyrosinase.
882	6	2.0	512	20	AAW14115	E. coli CadC prote	955	18	AAW38165	Human tyrosinase.
883	6	2.0	514	20	AAW83359	Streptococcus pneu	956	22	ABW64907	Arabidopsis thalia
884	6	2.0	514	23	ABW28786	Streptococcus poly	957	21	AAG10700	Arabidopsis thalia
885	6	2.0	514	23	ABW30078	Streptococcus poly	958	22	ABG22539	Novel human diagno
886	6	2.0	516	17	AAW94188	Pea ADPG-Phase BT2	959	23	ABW18327	Staphylococcus epi

960 6 2.0 565 22 ABB59416 Drosophila melanog  
 961 6 2.0 565 22 AAU36274 Pseudomonas aerugi  
 962 6 2.0 569 22 AAU34910 Enterococcus faeca  
 963 6 2.0 571 18 AAU55665 H. pylori ORF hp2e  
 964 6 2.0 571 21 AAG36716 Arabidopsis thalia  
 965 6 2.0 577 21 AAY71030 Ubiquitin-E. coli  
 966 6 2.0 578 22 ABG25010 Novel human diseno  
 967 6 2.0 579 19 AAU37876 Alcohol and/or ald  
 968 6 2.0 580 22 ABG09004 Novel human diago  
 969 6 2.0 581 21 AAG40001 Arabidopsis thalia  
 970 6 2.0 581 22 ABG62993 Drosophila melanog  
 971 6 2.0 582 22 AAU03896 G protein-coupled  
 972 6 2.0 582 23 ABP3476 Streptococcus poly  
 973 6 2.0 583 22 AAU37865 Streptococcus pneu  
 974 6 2.0 584 23 ABB55030 Lactococcus lactis  
 975 6 2.0 585 23 ABB48880 Listeria monocytog  
 976 6 2.0 588 23 ABP27904 Streptococcus poly  
 977 6 2.0 588 23 ABP29843 Streptococcus poly  
 978 6 2.0 590 10 AAP98204 Human muscarinic a  
 979 6 2.0 590 12 AAR12594 Antigenic portion  
 980 6 2.0 590 22 ABB56362 Non-endogenous hum  
 981 6 2.0 590 22 ABG27904 Novel human diago  
 982 6 2.0 590 22 AAM93244 Human polypeptide,  
 983 6 2.0 590 22 AAE01970 Human cholinergic  
 984 6 2.0 590 22 AAE01971 Human cholinergic  
 985 6 2.0 590 22 AAE01972 Human cholinergic  
 986 6 2.0 590 22 AAE01973 Human cholinergic  
 987 6 2.0 590 22 AAE01974 Human cholinergic  
 988 6 2.0 590 22 AAE01975 Human cholinergic  
 989 6 2.0 590 22 AAE01976 Human cholinergic  
 990 6 2.0 590 22 AAE01977 Human cholinergic  
 991 6 2.0 590 23 AAU98723 S. coelicolor Acca  
 992 6 2.0 590 23 AAU98724 S. coelicolor Acca  
 993 6 2.0 592 23 AAU83622 Human PRO protein,  
 994 6 2.0 593 22 AAB82668 Human tRNA synthet  
 995 6 2.0 593 23 AB07507 Human aminoacyl tr  
 996 6 2.0 596 14 AAR41433 PT-NANBH virus NS5  
 997 6 2.0 596 22 ABB65561 Drosophila melanog  
 998 6 2.0 597 22 ABG06708 Novel human diago  
 999 6 2.0 597 22 ABG26819 Novel human diago  
 1000 6 2.0 599 22 ABB71907 Drosophila melanog

## ALIGNMENTS

RESULT 1  
 AAY06965  
 ID AAY06965 standard; Protein; 293 AA.

AC AAY06965;  
 DT 05-JUL-1999 (first entry)  
 XX  
 DE E. canis P30-5 protein.  
 XX  
 KW Outer membrane protein; OMP; Ehrlichia chafeensis; E. canis; P30;  
 KW detection; dog.  
 XX  
 OS Ehrlichia canis.  
 XX  
 PN WO9913720-A1.  
 XX  
 PD 25-MAR-1999.  
 XX  
 PF 18-SEP-1998; 98WO-US19600.  
 XX  
 PR 19-SEP-1997; 97US-0059353.  
 XX  
 PA (OHIS ) UNIV OHIO STATE.  
 XX  
 PI Ohashi N, Rikihisa Y;  
 XX

DR WPI; 1999-254290/21.  
 DR N-PSDB; AAX34765.  
 XX  
 PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
 PT Ehrlichia canis  
 XX  
 PS Disclosure; Fig 25B; 55pp; English.  
 XX  
 CC The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and  
 CC consist of proteins shown in AAY06959-970. The proteins and genes are  
 CC used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX  
 SQ Sequence 293 AA;  
 Query Match 100.0%; Score 293; DB 20; Length 293;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-286;  
 Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MNNKLKFTIINTVLVCLLSLPNISSKAINNNNAKKYYGLYISGQYKPSVSFVSFVKET 60  
 DB 1 MNNKLKFTIINTVLVCLLSLPNISSKAINNNNAKKYYGLYISGQYKPSVSFVSFVKET 60  
 QY 61 NVITKNIALKKVDSTETKTDAVGISNPSNFTIPYTAVFQDNSVNFNGTGYTFAEGT 120  
 DB 61 NVITKNIALKKVDSTETKTDAVGISNPSNFTIPYTAVFQDNSVNFNGTGYTFAEGT 120  
 QY 121 RVEIGSYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPEKVSNSIFHTVMNDGLS 180  
 DB 121 RVEIGSYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPEKVSNSIFHTVMNDGLS 180  
 QY 181 IISVINVVCYDFSNLNLISPIYICGAGVDAIEPFDVLHFKFAYOSKLGIAYSLPNLSL 240  
 DB 181 IISVINVVCYDFSNLNLISPIYICGAGVDAIEPFDVLHFKFAYOSKLGIAYSLPNLSL 240  
 QY 241 FASLYYHKVMGNQKLNVLQHVLAELASIPKITSATVATLNIGYFGGEIGARLTF 293  
 DB 241 FASLYYHKVMGNQKLNVLQHVLAELASIPKITSATVATLNIGYFGGEIGARLTF 293

## RESULT 2

AAU96115  
 ID AAU96115 standard; Protein; 293 AA.

AC AAU96115;  
 DT 02-JUL-2002 (first entry)  
 XX  
 DE Ehrlichia canis p28-1.  
 XX  
 KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;  
 KW antibacterial.  
 XX  
 OS Ehrlichia canis.  
 XX  
 PN WO200222782-A2.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PF 12-SEP-2001; 2001WO-US28759.  
 XX  
 PR 12-SEP-2000; 2000US-0660597.  
 XX  
 PA (RERE-) RES DEV FOUND.  
 XX  
 PI Walker DH, Yu X, McBride JW;  
 XX  
 DR WPI; 2002-351882/38.  
 DR N-PSDB; ABK68875.  
 XX  
 PT New recombinant homologous 28 kilodalton immunodominant protein from



PT Ehrlichia canis, useful for treating Ehrlichia canis infections  
 PS Claim 16; Figure 13; 106pp; English.  
 XX  
 CC The invention relates to a recombinant homologous 28 kDa immunodominant  
 CC protein, P28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably  
 CC dispersed in a pharmaceutically acceptable carrier, is useful for  
 CC inhibiting E. canis infection in a subject. (I) is useful in the  
 CC development of vaccines and serodiagnostics that are particularly  
 CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118  
 CC represent the 28-kDa antigen amino acid sequences of the invention.  
 XX  
 SQ Sequence 293 AA;  
 Query Match 100.0%; Score 293; DB 23; Length 293;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-286;  
 Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MNKKLFTIINTVLVCLLSLPLNISSKAINNNAKKYGLYISGQYKPSVSVFSNFSVKET 60  
 DB 1 MNKKLFTIINTVLVCLLSLPLNISSKAINNNAKKYGLYISGQYKPSVSVFSNFSVKET 60  
 QY 61 NVITKNLIALKXVDSDIETKTDSVGSINSPNFTIPTYTAVFQDINSVNFNGTIGYTFARGT 120  
 DB 61 NVITKNLIALKXVDSDIETKTDSVGSINSPNFTIPTYTAVFQDINSVNFNGTIGYTFARGT 120  
 QY 121 RVEIGSYEEFDVKNPQGYTSLDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGLS 180  
 DB 121 RVEIGSYEEFDVKNPQGYTSLDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGLS 180  
 QY 181 IISVIVNVYDFSLNLSISPYICGAGVDAIEPDLVHLHKFAYOSKLGIAYSLSNLSL 240  
 DB 181 IISVIVNVYDFSLNLSISPYICGAGVDAIEPDLVHLHKFAYOSKLGIAYSLSNLSL 240  
 QY 241 FASLYYHKVMGNQFNKLVQHVLAELASIPKITSATVATNLNIGVFGGEIGARLTF 293  
 DB 241 FASLYYHKVMGNQFNKLVQHVLAELASIPKITSATVATNLNIGVFGGEIGARLTF 293  
 RESULT 3  
 AAU06949  
 ID AAU06949 standard; Protein; 280 AA.  
 XX  
 AC AAU06949;  
 XX  
 DT 05-JUL-1999 (first entry)  
 XX  
 DE E. chaffeensis OMP-1A protein.  
 XX  
 KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 KW detection; dog.  
 XX  
 OS Ehrlichia chaffeensis.  
 XX  
 PN WO9913720-A1.  
 XX  
 PD 25-MAR-1999.  
 XX  
 PF 18-SEP-1998; 98WO-US19600.  
 XX  
 PR 19-SEP-1997; 97US-0059353.  
 XX  
 PA (OHIS ) UNIV OHIO STATE.  
 XX  
 PI Ohashi N, Rikihisa Y;  
 XX  
 DR WPI; 1999-254290/21.  
 DR N-PSDB; AAX34749.  
 XX  
 PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
 PT Ehrlichia canis  
 XX  
 PS Claim 17; Fig 9B; 55pp; English.

XX The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 CC in AAU06943-958. The E. canis proteins form part of the P30 family and  
 CC consist of proteins shown in AAU06959-970. The proteins and genes are  
 CC used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX  
 SQ Sequence 280 AA;  
 Query Match 8.5%; Score 25; DB 20; Length 280;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-16;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 38 GLYISGQYKPSVSVFSNFSVKETNV 62  
 DB 25 GLYISGQYKPSVSVFSNFSVKETNV 49  
 RESULT 4  
 AAU73412  
 ID AAU73412 standard; Protein; 293 AA.  
 XX  
 AC AAU73412;  
 XX  
 DT 12-MAR-2002 (first entry)  
 XX  
 DE Ehrlichia chaffeensis outer membrane protein P28-13.  
 XX  
 KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.  
 XX  
 OS Ehrlichia chaffeensis.  
 XX  
 PN WO200183699-A2.  
 XX  
 PD 08-NOV-2001.  
 XX  
 PF 01-MAY-2001; 2001WO-US13997.  
 XX  
 PR 01-MAY-2000; 2000US-201035P.  
 XX  
 PA (RERE-) RES DEV FOUND.  
 XX  
 PI Walker DH, Yu X;  
 XX  
 DR WPI; 2002-066527/09.  
 XX  
 PT Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated  
 PT P28 useful as a vaccine against Ehrlichia chaffeensis -  
 XX  
 PS Claim 10; Figure 2; 97pp; English.  
 XX  
 CC The invention relates to isolated and purified 28-kDa outer membrane  
 CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins  
 CC are encoded by a 28kDa outer membrane protein multigene family. P28  
 CC proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28  
 CC is useful for transfecting a host cell. AAU73400-AAU73420 represent  
 CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.  
 XX  
 SQ Sequence 293 AA;  
 Query Match 8.5%; Score 25; DB 23; Length 293;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-16;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 38 GLYISGQYKPSVSVFSNFSVKETNV 62  
 DB 38 GLYISGQYKPSVSVFSNFSVKETNV 62  
 RESULT 5  
 AAU06967  
 ID AAU06967 standard; Protein; 292 AA.

```
XX AAY06967;
AC
XX
XX
DT 05-JUL-1999 (first entry)
XX
DE E. canis P30-7 protein.
XX
XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog.
XX
XX Ehrlichia canis.
OS
XX
XX WO9913720-A1.
PN
XX
XX 25-MAR-1999.
PD
XX
XX 18-SEP-1998; 98WO-US19600.
PF
XX
XX 19-SEP-1997; 97US-0059353.
PR
XX
XX (OHIS ) UNIV OHIO STATE.
PA
XX
XX Ohashi N, Rikihisa Y;
PI
XX
XX WPI: 1999-254290/21.
DR
XX N-PSDB; AAX34767.
DR
XX
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
PT
XX
XX Disclosure; Fig 27B; 55pp; English.
PS
XX
XX The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
XX Sequence 292 AA;

Query Match 8.2%; Score 24; DB 20; Length 292;
Best Local Similarity 100.0%; Pred. No. 1.6e-15;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GLYISQYKPSVSFNSFSVKETN 61
|||
DB 35 GLYISQYKPSVSFNSFSVKETN 58
|||

RESULT 6
AAU73410
ID AAU73410 standard; Protein; 298 AA.
XX
XX
XX AAU73410;
XX
XX
DT 12-MAR-2002 (first entry)
XX
DE Ehrlichia chaffeensis outer membrane protein P28-11.
XX
XX Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.
XX
XX Ehrlichia chaffeensis.
OS
XX
XX WO200183699-A2.
PN
XX
XX 08-NOV-2001.
PD
XX
XX 01-MAY-2001; 2001WO-US13997.
PF
XX
XX 01-MAY-2000; 2000US-201035P.
PR
XX
XX (RERE-) RES DEV FOUND.
PA
```

```
XX Walker DH, Yu X;
PI
XX
XX WPI: 2002-066527/09.
DR
XX
XX Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated
PT P28 useful as a vaccine against Ehrlichia chaffeensis
XX
XX Claim 10; Figure 2; 97pp; English.
PS
XX
XX The invention relates to isolated and purified 28-kDa outer membrane
CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins
CC are encoded by a 28kDa outer membrane protein multigene family. P28
CC proteins are useful as a vaccine against E. chaffeensis. DNA encoding P28
CC is useful for transfecting a host cell. AAU73400-AAU73420 represent
CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.
XX
XX Sequence 298 AA;
SQ

Query Match 7.5%; Score 22; DB 23; Length 298;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 YISQYKPSVSFNSFSVKETN 61
|||
DB 39 YISQYKPSVSFNSFSVKETN 60
|||

RESULT 7
AAU06958
ID AAU06958 standard; Protein; 271 AA.
XX
XX
XX AAU06958;
AC
XX
XX 05-JUL-1999 (first entry)
DT
XX
XX E. chaffeensis OMP-1Z protein.
DE
XX
XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog.
XX
XX Ehrlichia chaffeensis.
OS
XX
XX WO9913720-A1.
PN
XX
XX 25-MAR-1999.
PD
XX
XX 18-SEP-1998; 98WO-US19600.
PF
XX
XX 19-SEP-1997; 97US-0059353.
PR
XX
XX (OHIS ) UNIV OHIO STATE.
PA
XX
XX Ohashi N, Rikihisa Y;
PI
XX
XX WPI: 1999-254290/21.
DR
XX N-PSDB; AAX34758.
DR
XX
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
PT
XX
XX Disclosure; Fig 18B; 55pp; English.
PS
XX
XX The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
XX Sequence 271 AA;
SQ

Query Match 5.8%; Score 17; DB 20; Length 271;
```

Best Local Similarity 100.0%; Pred. No. 1.6e-08; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 VMGNQFKNLNVQHVAEL 265  
 |||||  
 Db 227 VMGNQFKNLNVQHVAEL 243

RESULT 8  
 AAU73411  
 ID AAU73411 standard; Protein; 300 AA.  
 XX AC AAU73411;  
 XX DT 12-MAR-2002 (first entry)  
 XX DE Ehrlichia chaffeensis outer membrane protein P28-12.  
 XX KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.  
 XX OS Ehrlichia chaffeensis.  
 XX PN WO200183699-A2.  
 XX PD 08-NOV-2001.  
 XX PF 01-MAY-2001; 2001WO-US13997.  
 XX PR 01-MAY-2000; 2000US-201035P.  
 XX PA (RERE-) RES DEV FOUND.  
 XX PI Walker DH, Yu X;  
 XX DR WPI; 2002-066527/09.  
 XX PT Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated P28 useful as a vaccine against Ehrlichia chaffeensis -  
 XX PS Claim 10; Figure 2; 97pp; English.  
 XX SS The invention relates to isolated and purified 28-kDa outer membrane proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins are encoded by a 28kDa outer membrane protein multigene family. P28 proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28 is useful for transfecting a host cell. AAU73400-AAU73420 represent Ehrlichia chaffeensis P28 outer membrane proteins of the invention.  
 XX SQ Sequence 300 AA;

Query Match 5.8%; Score 17; DB 23; Length 300;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-08;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 VMGNQFKNLNVQHVAEL 265  
 |||||  
 Db 256 VMGNQFKNLNVQHVAEL 272

RESULT 9  
 AAU06951  
 ID AAU06951 standard; Protein; 164 AA.  
 XX AC AAU06951;  
 XX DT 05-JUL-1999 (first entry)  
 XX DE E. chaffeensis OMP-1S protein.  
 XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 XX detection; dog.  
 XX OS Ehrlichia chaffeensis.

XX WO9913720-A1.  
 PN 25-MAR-1999.  
 PD 18-SEP-1998; 98WO-US19600.  
 PF 19-SEP-1997; 97US-0059353.  
 PR (OHIS ) UNIV OHIO STATE.  
 PA Ohashi N, Rikihisa Y;  
 PI WPI; 1999-254290/21.  
 XX DR N-ESDB; AAX34751.  
 XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and Ehrlichia canis  
 XX PS Disclosure; Fig 11B; 55pp; English.  
 XX CC The invention provides isolated outer membrane proteins (OMP) from Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part of the OMP family and consist of proteins OMP-1, -1(B to Z) shown in AAU06943-958. The E. canis proteins form part of the P30 family and consist of proteins shown in AAU06959-970. The proteins and genes are used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX SQ Sequence 164 AA;

Query Match 5.5%; Score 16; DB 20; Length 164;  
 Best Local Similarity 100.0%; Pred. No. 1e-07;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GLYISGOYKPSVSVFS 53  
 |||||  
 Db 37 GLYISGOYKPSVSVFS 52

RESULT 10  
 AAU73409  
 ID AAU73409 standard; Protein; 291 AA.  
 XX AC AAU73409;  
 XX DT 12-MAR-2002 (first entry)  
 XX DE Ehrlichia chaffeensis outer membrane protein P28-10.  
 XX KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.  
 XX OS Ehrlichia chaffeensis.  
 XX PN WO200183699-A2.  
 XX PD 08-NOV-2001.  
 XX PF 01-MAY-2001; 2001WO-US13997.  
 XX PR 01-MAY-2000; 2000US-201035P.  
 XX PA (RERE-) RES DEV FOUND.  
 XX PI Walker DH, Yu X;  
 XX DR WPI; 2002-066527/09.  
 XX PT Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated P28 useful as a vaccine against Ehrlichia chaffeensis -  
 XX PS Claim 10; Figure 2; 97pp; English.  
 XX CC The invention relates to isolated and purified 28-kDa outer membrane

CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins  
 CC are encoded by a 28kDa outer membrane protein multigene family. P28  
 CC proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28  
 CC is useful for transfecting a host cell. AAU73400-AAU73420 represent  
 CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.  
 XX  
 SQ Sequence 291 AA;

Query Match 5.1%; Score 16; DB 23; Length 291;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GLYISQYKPSVSFVS 53  
 |||||  
 DB 37 GLYISQYKPSVSFVS 52

RESULT 11  
 ID AAY06956 standard; Protein; 275 AA.  
 XX  
 AC AAY06956;  
 XX  
 DT 05-JUL-1999 (first entry)  
 XX  
 DE E. chaffeensis OMP-1X protein.  
 XX  
 KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 KW detection; dog.  
 XX  
 OS Ehrlichia chaffeensis.  
 XX  
 PN WO9913720-A1.  
 XX  
 PD 25-MAR-1999.  
 XX  
 PF 18-SEP-1998; 98WO-US19600.  
 XX  
 PR 19-SEP-1997; 97US-0059353.  
 XX  
 PA (OHIS ) UNIV OHIO STATE.  
 XX  
 PI Ohaehi N, Rikihisa Y;  
 XX  
 DR WPI; 1999-254290/21.  
 DR N-PSDB; AAX34756.  
 XX  
 PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
 XX Ehrlichia canis  
 PS Disclosure; Fig 16B; 55pp; English.  
 XX  
 CC The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -(B to Z) shown  
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and  
 CC consist of proteins shown in AAY06959-970. The proteins and genes are  
 CC used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX  
 SQ Sequence 275 AA;

Query Match 5.1%; Score 15; DB 20; Length 275;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 274 AVATLNIGYFGGEIG 288  
 |||||  
 DB 256 AVATLNIGYFGGEIG 270

RESULT 12  
 ID AAU73407 standard; Protein; 275 AA.  
 XX

XX AAU73407;  
 AC  
 XX 12-MAR-2002 (first entry)  
 DT  
 XX Ehrlichia chaffeensis outer membrane protein P28-8.  
 DE  
 XX Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.  
 KW  
 XX Ehrlichia chaffeensis.  
 OS  
 XX WO200183699-A2.  
 PN  
 XX 08-NOV-2001.  
 PD  
 XX 01-MAY-2001; 2001WO-US13997.  
 PF  
 XX 01-MAY-2000; 2000US-201035P.  
 PR  
 XX (RERE-) RES DEV FOUND.  
 PA  
 XX Walker DH, Yu X;  
 PI  
 XX WPI; 2002-066527/09.  
 DR  
 XX Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated  
 PT P28 useful as a vaccine against Ehrlichia chaffeensis -  
 PT  
 XX Claim 10; Figure 2; 97pp; English.  
 PS  
 XX The invention relates to isolated and purified 28-kDa outer membrane  
 CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins  
 CC are encoded by a 28kDa outer membrane protein multigene family. P28  
 CC proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28  
 CC is useful for transfecting a host cell. AAU73400-AAU73420 represent  
 CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.  
 XX  
 SQ Sequence 275 AA;

Query Match 5.1%; Score 15; DB 23; Length 275;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 274 AVATLNIGYFGGEIG 288  
 |||||  
 DB 256 AVATLNIGYFGGEIG 270

RESULT 13  
 ID AAY06969 standard; Protein; 161 AA.  
 XX  
 AC AAY06969;  
 XX  
 DT 05-JUL-1999 (first entry)  
 XX  
 DE E. canis P30-9 protein.  
 DE  
 KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 KW detection; dog.  
 KW  
 XX Ehrlichia canis.  
 OS  
 XX WO9913720-A1.  
 PN  
 XX 25-MAR-1999.  
 PD  
 XX 18-SEP-1998; 98WO-US19600.  
 PF  
 XX 19-SEP-1997; 97US-0059353.  
 PR  
 XX (OHIS ) UNIV OHIO STATE.  
 PA  
 XX



XX WO9913720-A1.  
 XX PN  
 XX 25-MAR-1999.  
 XX PD  
 XX 18-SEP-1998; 98WO-US19600.  
 XX PF  
 XX 19-SEP-1997; 97US-0059353.  
 XX PR  
 XX (OHIS ) UNIV OHIO STATE.  
 XX PA  
 XX Ohashi N, Rikihisa Y;  
 XX PI  
 XX WPI; 1999-254290/21.  
 XX DR  
 XX N-PSDB; AAX34755.  
 XX DR  
 XX Novel outer membrane proteins from Ehrlichia chaffeensis and  
 XX PT  
 XX Ehrlichia canis  
 XX PT  
 XX Disclosure; Fig 15B; 55pp; English.  
 XX PS  
 XX The invention provides isolated outer membrane proteins (OMP) from  
 XX CC  
 XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 XX CC  
 XX of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 XX CC  
 XX in AAY06943-958. The E. canis proteins form part of the P30 family and  
 XX CC  
 XX consist of proteins shown in AAY06959-970. The proteins and genes are  
 XX CC  
 XX used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX CC  
 XX Sequence 283 AA;  
 XX SQ

Query Match 3.8%; Score 11; DB 20; Length 283;  
 Best Local Similarity 100.0%; Pred. No. 0.018;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 SYEEFDVKNPG 137  
 |||||  
 DB 122 SYEEFDVKNPG 132

RESULT 17  
 AAU73406  
 ID AAU73406 standard; Protein; 283 AA.  
 XX AC  
 XX AAU73406;  
 XX DT  
 XX 12-MAR-2002 (first entry)  
 XX DE  
 XX Ehrlichia chaffeensis outer membrane protein P28-7.  
 XX KW  
 XX Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.  
 XX OS  
 XX Ehrlichia chaffeensis.  
 XX PN  
 XX WO200183699-A2.  
 XX XX  
 XX 08-NOV-2001.  
 XX PD  
 XX 01-MAY-2001; 2001WO-US13997.  
 XX PF  
 XX 01-MAY-2000; 2000US-201035P.  
 XX PR  
 XX (RERE-) RES DEV FOUND.  
 XX PA  
 XX Walker DH, Yu X;  
 XX PI  
 XX WPI; 2002-066527/09.  
 XX DR  
 XX Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated  
 XX PT  
 XX P28 useful as a vaccine against Ehrlichia chaffeensis  
 XX PS  
 XX Claim 10; Figure 2; 97pp; English.  
 XX CC  
 XX The invention relates to isolated and purified 28-kDa outer membrane

CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins  
 CC are encoded by a 28kDa outer membrane protein multigene family. P28  
 CC proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28  
 CC is useful for transfecting a host cell. AAU73400-AAU73420 represent  
 CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.  
 XX CC  
 XX SQ Sequence 283 AA;  
 XX

Query Match 3.8%; Score 11; DB 23; Length 283;  
 Best Local Similarity 100.0%; Pred. No. 0.018;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 SYEEFDVKNPG 137  
 |||||  
 DB 122 SYEEFDVKNPG 132

RESULT 18  
 AAY06953  
 ID AAY06953 standard; Protein; 295 AA.  
 XX AC  
 XX AAY06953;  
 XX DT  
 XX 05-JUL-1999 (first entry)  
 XX DE  
 XX E. chaffeensis OMP-1U protein.  
 XX KW  
 XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 XX KW  
 XX detection; dog.  
 XX OS  
 XX Ehrlichia chaffeensis.  
 XX PN  
 XX WO9913720-A1.  
 XX PD  
 XX 25-MAR-1999.  
 XX PF  
 XX 18-SEP-1998; 98WO-US19600.  
 XX XX  
 XX 19-SEP-1997; 97US-0059353.  
 XX PR  
 XX (OHIS ) UNIV OHIO STATE.  
 XX PA  
 XX Ohashi N, Rikihisa Y;  
 XX PI  
 XX WPI; 1999-254290/21.  
 XX DR  
 XX N-PSDB; AAX34753.  
 XX DR  
 XX Novel outer membrane proteins from Ehrlichia chaffeensis and  
 XX PT  
 XX Ehrlichia canis  
 XX PS  
 XX Disclosure; Fig 13B; 55pp; English.  
 XX OS  
 XX The invention provides isolated outer membrane proteins (OMP) from  
 XX CC  
 XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 XX CC  
 XX of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 XX CC  
 XX in AAY06943-958. The E. canis proteins form part of the P30 family and  
 XX CC  
 XX consist of proteins shown in AAY06959-970. The proteins and genes are  
 XX CC  
 XX used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX CC  
 XX Sequence 295 AA;  
 XX SQ

Query Match 3.8%; Score 11; DB 20; Length 295;  
 Best Local Similarity 100.0%; Pred. No. 0.018;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GLYISGQYKPS 48  
 |||||  
 DB 38 GLYISGQYKPS 48

RESULT 19  
 AAU73404  
 ID AAU73404 standard; Protein; 295 AA.

XX AC AAU73404;  
 XX 12-MAR-2002 (first entry)  
 XX Ehrlichia chaffeensis outer membrane protein P28-5.  
 XX Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.  
 XX OS Ehrlichia chaffeensis.  
 XX WO200183699-A2.  
 XX PD 08-NOV-2001.  
 XX PF 01-MAY-2001; 2001WO-US13997.  
 XX PR 01-MAY-2000; 2000US-201035P.  
 XX PA (RERE-) RES DEV FOUND.  
 XX PI Walker DH, Yu X;  
 XX DR WPI; 2002-066527/09.  
 XX PT Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated  
 XX P28 useful as a vaccine against Ehrlichia chaffeensis -  
 XX PS Claim 10; Figure 2; 97pp; English.  
 XX CC The invention relates to isolated and purified 28-kDa outer membrane  
 XX proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins  
 XX are encoded by a 28kDa outer membrane protein multigene family. P28  
 XX proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28  
 XX is useful for transfecting a host cell. AAU73400-AAU73420 represent  
 XX Ehrlichia chaffeensis P28 outer membrane proteins of the invention.  
 XX SQ Sequence 295 AA;  
 Query Match 3.8%; Score 11; DB 23; Length 295;  
 Best Local Similarity 100.0%; Pred.No. 0.018;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 38 GLYISGQYKPS 48  
 Db 38 GLYISGQYKPS 48  
 RESULT 20  
 AAU06968  
 ID AAU06968 standard; Protein; 299 AA.  
 XX AC AAU06968;  
 XX DT 05-JUL-1999 (first entry)  
 XX DE E. canis P30-8 protein.  
 XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 XX detection; dog.  
 XX OS Ehrlichia canis.  
 XX WO9913720-A1.  
 XX 25-MAR-1999.  
 XX PF 18-SEP-1998; 98WO-US19600.  
 XX PR 19-SEP-1997; 97US-0059353.  
 XX PA (OHIS ) UNIV OHIO STATE.  
 XX PI Ohashi N, Rikihisa Y;  
 XX DR WPI; 1999-254290/21.  
 XX DR N-PSDB; AAX34757.  
 XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
 XX Ehrlichia canis.  
 XX PS Disclosure; Fig 17B; 55pp; English.  
 XX CC The invention provides isolated outer membrane proteins (OMP) from  
 XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 XX of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 XX in AAU06943-958. The E. canis proteins form part of the P30 family and  
 XX consist of proteins shown in AAU06959-970. The proteins and genes are  
 XX used to detect E. chaffeensis in patients and E. canis in dogs.

PI Ohashi N, Rikihisa Y;  
 XX WPI; 1999-254290/21.  
 XX DR N-PSDB; AAX34768.  
 XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
 XX Ehrlichia canis  
 XX PS Disclosure; Fig 28B; 55pp; English.  
 XX CC The invention provides isolated outer membrane proteins (OMP) from  
 XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 XX of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 XX in AAU06943-958. The E. canis proteins form part of the P30 family and  
 XX consist of proteins shown in AAU06959-970. The proteins and genes are  
 XX used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX SQ Sequence 299 AA;  
 Query Match 3.8%; Score 11; DB 20; Length 299;  
 Best Local Similarity 100.0%; Pred.No. 0.019;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 269 PKITSAVATLN 279  
 Db 275 PKITSAVATLN 285  
 RESULT 21  
 AAU06957  
 ID AAU06957 standard; Protein; 285 AA.  
 XX AC AAU06957;  
 XX DT 05-JUL-1999 (first entry)  
 XX DE E. chaffeensis OMP-1Y protein.  
 XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 XX detection; dog.  
 XX OS Ehrlichia chaffeensis.  
 XX WO9913720-A1.  
 XX 25-MAR-1999.  
 XX PF 18-SEP-1998; 98WO-US19600.  
 XX PR 19-SEP-1997; 97US-0059353.  
 XX PA (OHIS ) UNIV OHIO STATE.  
 XX PI Ohashi N, Rikihisa Y;  
 XX DR WPI; 1999-254290/21.  
 XX DR N-PSDB; AAX34757.  
 XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
 XX Ehrlichia canis  
 XX PS Disclosure; Fig 17B; 55pp; English.  
 XX CC The invention provides isolated outer membrane proteins (OMP) from  
 XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 XX of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 XX in AAU06943-958. The E. canis proteins form part of the P30 family and  
 XX consist of proteins shown in AAU06959-970. The proteins and genes are  
 XX used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX SQ Sequence 285 AA;  
 Query Match 3.4%; Score 10; DB 20; Length 285;

```

Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 SYEEFDVKNP 136
DB 123 SYEEFDVKNP 132

RESULT 22
AAU73408
ID AAU73408 standard; Protein; 285 AA.
XX
AC AAU73408;
XX
DT 12-MAR-2002 (first entry)
XX
DE Ehrlichia chaffeensis outer membrane protein P28-9.
XX
KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.
XX
OS Ehrlichia chaffeensis.
XX
PN WO200183699-A2.
XX
PD 08-NOV-2001.
XX
PF 01-MAY-2001; 2001WO-US13997.
XX
PR 01-MAY-2000; 2000US-201035P.
XX
PA (RERE-) RES DEV FOUND.
XX
PI Walker DH, Yu X;
XX
DR WPI; 2002-066527/09.
XX
PT Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated
PT P28 useful as a vaccine against Ehrlichia chaffeensis
XX
PS Claim 10; Figure 2; 97pp; English.
XX
CC The invention relates to isolated and purified 28-kDa outer membrane
CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins
CC are encoded by a 28kba outer membrane protein multigene family. P28
CC proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28
CC is useful for transfecting a host cell. AAU73400-AAU73420 represent
CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.
XX
SQ Sequence 285 AA;

Query Match 3.4%; Score 10; DB 23; Length 285;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 SYEEFDVKNP 136
DB 123 SYEEFDVKNP 132

RESULT 23
AAU06966
ID AAU06966 standard; Protein; 289 AA.
XX
AC AAU06966;
XX
DT 05-JUL-1999 (first entry)
XX
DE E. canis P30-6 protein.
XX
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog.
XX
OS Ehrlichia canis.

```

```

XX WO9913720-A1.
PN
XX 25-MAR-1999.
PD
XX 18-SEP-1998; 98WO-US19600.
PF
XX 19-SEP-1997; 97US-0059353.
PR
XX (OHIS ) UNIV OHIO STATE.
PA
XX Ohashi N, Rikihisa Y;
PI
XX WPI; 1999-254290/21.
DR
XX N-PSDB; AAX34766.
DR
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
PT Disclosure; Fig 26B; 55pp; English.
PS
XX The invention provides isolated outer membrane proteins (OMP) from
XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAU06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAU06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
SQ Sequence 289 AA;

Query Match 3.4%; Score 10; DB 20; Length 289;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 PSVSVSFNFS 56
DB 44 PSVSVSFNFS 53

RESULT 24
AAU06952
ID AAU06952 standard; Protein; 131 AA.
XX
AC AAU06952;
XX
DT 05-JUL-1999 (first entry)
XX
DE E. chaffeensis OMP-1T protein.
XX
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog.
XX
OS Ehrlichia chaffeensis.
XX
PN WO9913720-A1.
PD
XX 25-MAR-1999.
PF
XX 18-SEP-1998; 98WO-US19600.
PR
XX 19-SEP-1997; 97US-0059353.
XX
XX (OHIS ) UNIV OHIO STATE.
PA
XX Ohashi N, Rikihisa Y;
PI
XX WPI; 1999-254290/21.
DR
XX N-PSDB; AAX34752.
XX
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
PS Disclosure; Fig 12B; 55pp; English.

```



XX The invention provides isolated outer membrane proteins (OMP) from  
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
CC in AAU06943-958. The E. canis proteins form part of the p30 family and  
CC consist of proteins shown in AAU06959-970. The proteins and genes are  
CC used to detect E. chaffeensis in patients and E. canis in dogs.  
XX  
SQ Sequence 131 AA;  
Query Match 2.7%; Score 8; DB 20; Length 131;  
Best Local Similarity 100.0%; Pred. No. 9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 245 YYHKVMGN 252  
| | | | | | | |  
Db 82 YYHKVMGN 89  
RESULT 25  
AAU73403  
ID AAU73403 standard; Protein; 272 AA.  
XX  
AC AAU73403;  
XX  
DT 12-MAR-2002 (first entry)  
XX  
DE Ehrlichia chaffeensis outer membrane protein P28-4.  
XX  
KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.  
XX  
OS Ehrlichia chaffeensis.  
XX  
XX WO200183699-A2.  
XX  
XX 08-NOV-2001.  
XX  
PF -01-MAY-2001; 2001WO-US13997.  
XX  
PR 01-MAY-2000; 2000US-201035P.  
XX  
PA (RERE-) RES DEV FOUND.  
XX  
PI Walker DH, Yu X;  
XX  
DR WPI; 2002-066527/09.  
XX  
PT Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated  
PT P28 useful as a vaccine against Ehrlichia chaffeensis -  
XX  
PS Claim 10; Figure 2; 97pp; English.  
XX  
CC The invention relates to isolated and purified 28-kDa outer membrane  
CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins  
CC are encoded by a 28kDa outer membrane protein multigene family. P28  
CC proteins are useful as a vaccine against E. chaffeensis. DNA encoding P28  
CC is useful for transfecting a host cell. AAU73400-AAU73420 represent  
CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.  
XX  
SQ Sequence 272 AA;  
Query Match 2.7%; Score 8; DB 23; Length 272;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 245 YYHKVMGN 252  
| | | | | | | |  
Db 223 YYHKVMGN 230  
RESULT 26  
AAU96116  
ID AAU96116 standard; Protein; 280 AA.  
XX  
XX The invention provides isolated outer membrane proteins (OMP) from  
XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
XX of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
XX in AAU06943-958. The E. canis proteins form part of the p30 family and  
XX consist of proteins shown in AAU06959-970. The proteins and genes are  
XX used to detect E. chaffeensis in patients and E. canis in dogs.  
XX  
SQ Sequence 131 AA;  
Query Match 2.7%; Score 8; DB 20; Length 131;  
Best Local Similarity 100.0%; Pred. No. 9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 245 YYHKVMGN 252  
| | | | | | | |  
Db 82 YYHKVMGN 89  
RESULT 25  
AAU73403  
ID AAU73403 standard; Protein; 272 AA.  
XX  
AC AAU73403;  
XX  
DT 12-MAR-2002 (first entry)  
XX  
DE Ehrlichia chaffeensis outer membrane protein P28-4.  
XX  
KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.  
XX  
OS Ehrlichia chaffeensis.  
XX  
XX WO200183699-A2.  
XX  
XX 08-NOV-2001.  
XX  
PF -01-MAY-2001; 2001WO-US13997.  
XX  
PR 01-MAY-2000; 2000US-201035P.  
XX  
PA (RERE-) RES DEV FOUND.  
XX  
PI Walker DH, Yu X;  
XX  
DR WPI; 2002-066527/09.  
XX  
PT Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated  
PT P28 useful as a vaccine against Ehrlichia chaffeensis -  
XX  
PS Claim 10; Figure 2; 97pp; English.  
XX  
CC The invention relates to isolated and purified 28-kDa outer membrane  
CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins  
CC are encoded by a 28kDa outer membrane protein multigene family. P28  
CC proteins are useful as a vaccine against E. chaffeensis. DNA encoding P28  
CC is useful for transfecting a host cell. AAU73400-AAU73420 represent  
CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.  
XX  
SQ Sequence 272 AA;  
Query Match 2.7%; Score 8; DB 23; Length 272;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 245 YYHKVMGN 252  
| | | | | | | |  
Db 223 YYHKVMGN 230  
RESULT 26  
AAU96116  
ID AAU96116 standard; Protein; 280 AA.  
XX  
XX The invention provides isolated outer membrane proteins (OMP) from  
XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
XX of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
XX in AAU06943-958. The E. canis proteins form part of the p30 family and  
XX consist of proteins shown in AAU06959-970. The proteins and genes are  
XX used to detect E. chaffeensis in patients and E. canis in dogs.  
XX  
SQ Sequence 131 AA;  
Query Match 2.7%; Score 8; DB 20; Length 131;  
Best Local Similarity 100.0%; Pred. No. 9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 245 YYHKVMGN 252  
| | | | | | | |  
Db 82 YYHKVMGN 89  
RESULT 25  
AAU73403  
ID AAU73403 standard; Protein; 272 AA.  
XX  
AC AAU73403;  
XX  
DT 12-MAR-2002 (first entry)  
XX  
DE Ehrlichia chaffeensis outer membrane protein P28-4.  
XX  
KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.  
XX  
OS Ehrlichia chaffeensis.  
XX  
XX WO200183699-A2.  
XX  
XX 08-NOV-2001.  
XX  
PF -01-MAY-2001; 2001WO-US13997.  
XX  
PR 01-MAY-2000; 2000US-201035P.  
XX  
PA (RERE-) RES DEV FOUND.  
XX  
PI Walker DH, Yu X, McBride JW;  
XX  
DR WPI; 2002-351882/38.  
XX  
DR N-PSDB; ABK68876.  
XX  
PT New recombinant homologous 28 kilodalton immunodominant protein from  
PT Ehrlichia canis, useful for treating Ehrlichia canis infections -  
XX  
PS Claim 16; Figure 14; 106pp; English.  
XX  
CC The invention relates to a recombinant homologous 28 kDa immunodominant  
CC protein, P28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably  
CC dispersed in a pharmaceutically acceptable carrier, is useful for  
CC inhibiting E. canis infection in a subject. (I) is useful in the  
CC development of vaccines and serodiagnostics that are particularly  
CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118  
CC represent the 28-kDa antigen amino acid sequences of the invention.  
XX  
SQ Sequence 280 AA;  
Query Match 2.7%; Score 8; DB 23; Length 280;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 281 GYFGGEIG 288  
| | | | | | | |  
Db 268 GYFGGEIG 275  
RESULT 27  
ABP10881  
ID ABP10881 standard; Protein; 65 AA.  
XX  
AC ABP10881;  
XX  
DT 24-JUN-2002 (first entry)  
XX  
DE Human ORFX protein sequence SEQ ID NO:21744.  
XX  
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KW myasthenia gravis.  
XX  
OS Homo sapiens.  
XX  
PN WO200192523-A2.

XX AAU96116;  
XX AC  
XX 02-JUL-2002 (first entry)  
DT  
XX Ehrlichia canis p28-2.  
DE  
XX Ehrlichia canis infection; vaccine; serodiagnostic; p28;  
KW antibacterial.  
KW  
XX Ehrlichia canis.  
OS  
XX WO200222782-A2..  
PN  
XX 21-MAR-2002.  
PD  
XX 12-SEP-2001; 2001WO-US28759.  
PP  
XX 12-SEP-2000; 2000US-0660587.  
PR  
XX (RERE-) RES DEV FOUND.  
PA  
XX Walker DH, Yu X, McBride JW;  
PI  
XX WPI; 2002-351882/38.  
DR  
XX N-PSDB; ABK68876.  
DR  
XX New recombinant homologous 28 kilodalton immunodominant protein from  
PT Ehrlichia canis, useful for treating Ehrlichia canis infections -  
PT  
XX Claim 16; Figure 14; 106pp; English.  
PS  
XX The invention relates to a recombinant homologous 28 kDa immunodominant  
CC protein, P28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably  
CC dispersed in a pharmaceutically acceptable carrier, is useful for  
CC inhibiting E. canis infection in a subject. (I) is useful in the  
CC development of vaccines and serodiagnostics that are particularly  
CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118  
CC represent the 28-kDa antigen amino acid sequences of the invention.  
XX  
SQ Sequence 280 AA;  
Query Match 2.7%; Score 8; DB 23; Length 280;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 281 GYFGGEIG 288  
| | | | | | | |  
Db 268 GYFGGEIG 275  
RESULT 27  
ABP10881  
ID ABP10881 standard; Protein; 65 AA.  
XX  
AC ABP10881;  
XX  
DT 24-JUN-2002 (first entry)  
XX  
DE Human ORFX protein sequence SEQ ID NO:21744.  
XX  
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KW myasthenia gravis.  
XX  
OS Homo sapiens.  
XX  
PN WO200192523-A2.



```

XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00669.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483447/52.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human fetal liver -
XX PS Claim 27; SEQ ID NO 35977; 639pp + sequence listing; English.
XX CC The invention relates to a single exon nucleic acid probe for
XX CC measuring human gene expression in a sample derived from human foetal
XX CC liver. The single exon nucleic acid probes may be used for predicting,
XX CC measuring and displaying gene expression in samples derived from human
XX CC fetal liver. The present sequence is a peptide encoded by a single exon
XX CC nucleic acid probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 101 AA;

Query Match 2.4%; Score 7; DB 22; Length 101;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 QDNSVNF 108
Db 61 QDNSVNF 67
|||||
|||||

RESULT 30'
AAM64250
ID AAM64250 standard; Protein; 101 AA.
XX AC AAM64250;
XX DT 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 36355.
XX KW Human; brain expressed exon; gene expression analysis; probe;
XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX KW epilepsy; cancer.
XX OS Homo sapiens.
XX FN WO200157275-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00667.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488900/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human bone marrow -
XX PS Example 4; SEQ ID NO: 37380; 658pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid

```

---

```

PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483446/52.
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains -
XX PS Example 4; SEQ ID NO: 36355; 650pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC brain. They can be used to measure gene expression in brain cell samples,
XX CC which may enable the diagnosis and improved treatment of nervous system
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX CC epilepsy and cancers. The present sequence is a protein encoded by one of
XX CC the probes of the invention.
XX SQ Sequence 101 AA;

Query Match 2.4%; Score 7; DB 22; Length 101;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 QDNSVNF 108
Db 61 QDNSVNF 67
|||||
|||||

RESULT 31
AAM77074
ID AAM77074 standard; Protein; 101 AA.
XX AC AAM77074;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 37380.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX FN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00668.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488900/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human bone marrow -
XX PS Example 4; SEQ ID NO: 37380; 658pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid

```

CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
CC protein encoded by one of the probes of the invention.

XX SQ Sequence 101 AA;

Query Match 2.4%; Score 7; DB 22; Length 101;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 QDQSVNF 108

Db 61 QDQSVNF 67

RESULT 32

AA37196  
ID AA37196 standard; Protein; 101 AA.

XX AC AA37196;

XX DT 17-OCT-2001 (first entry)

XX DE Peptide #11233 encoded by probe for measuring placental gene expression.

XX KW Probe; microarray; human; placenta; antenatal diagnosis;  
XX KW genetic disorder.

XX OS Homo sapiens.

XX PN WO200157272-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00663.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488897/53.

XX PT Human genome-derived single exon nucleic acid probes useful for  
XX PT analyzing gene expression in human placenta -

XX PS Claim 27; SEQ ID No 37465; 654pp; English.

XX CC The present invention relates to single exon nucleic acid probes (SENPs;  
XX CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one  
XX CC such probe. The probes are useful for producing a microarray for  
XX CC predicting, measuring and displaying gene expression in samples derived  
XX CC from human placenta. The probes are useful for antenatal diagnosis of  
XX CC human genetic disorders.

XX SQ Sequence 101 AA;

Query Match 2.4%; Score 7; DB 22; Length 101;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 QDQSVNF 108

Db 61 QDQSVNF 67

RESULT 33

AA21513

ID AA21513 standard; Protein; 115 AA.

XX AC AA21513;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 24095.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127462.

XX PR 06-APR-1999; 99US-0128234.

XX PR 08-APR-1999; 99US-0128714.

XX PR 16-APR-1999; 99US-0129845.

XX PR 19-APR-1999; 99US-0130077.

XX PR 21-APR-1999; 99US-0130449.

XX PR 23-APR-1999; 99US-0130510.

XX PR 28-APR-1999; 99US-0130891.

XX PR 30-APR-1999; 99US-0131449.

XX PR 30-APR-1999; 99US-0132048.

XX PR 04-MAY-1999; 99US-0132407.

XX PR 05-MAY-1999; 99US-0132484.

XX PR 06-MAY-1999; 99US-0132485.

XX PR 06-MAY-1999; 99US-0132486.

XX PR 07-MAY-1999; 99US-0132487.

XX PR 11-MAY-1999; 99US-0132863.

XX PR 14-MAY-1999; 99US-0134218.

XX PR 14-MAY-1999; 99US-0134219.

XX PR 14-MAY-1999; 99US-0134221.

XX PR 18-MAY-1999; 99US-0134370.

XX PR 19-MAY-1999; 99US-0134768.

XX PR 20-MAY-1999; 99US-0134941.

XX PR 21-MAY-1999; 99US-0135124.

XX PR 21-MAY-1999; 99US-0135353.

XX PR 24-MAY-1999; 99US-0135629.

XX PR 25-MAY-1999; 99US-0136021.

XX PR 27-MAY-1999; 99US-0136392.

XX PR 28-MAY-1999; 99US-0136782.

XX PR 01-JUN-1999; 99US-0137222.

XX PR 03-JUN-1999; 99US-0137528.

XX PR 04-JUN-1999; 99US-0137502.

XX PR 07-JUN-1999; 99US-0137724.

XX PR 08-JUN-1999; 99US-0138094.

XX PR 10-JUN-1999; 99US-0138540.

XX PR 10-JUN-1999; 99US-0138847.

XX PR 14-JUN-1999; 99US-0139119.

XX PR 16-JUN-1999; 99US-0139452.

XX PR 16-JUN-1999; 99US-0139453.

XX PR 17-JUN-1999; 99US-0139492.

XX PR 18-JUN-1999; 99US-0139454.

XX PR 18-JUN-1999; 99US-0139455.

XX PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.

PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158212.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 2.4%; Score 7; DB 21; Length 115;  
Best local Similarity 100.0%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 NTVLVCL 17  
Db 65 NTVLVCL 71

RESULT 34  
ABB70338  
ID .ABB70338 standard; Protein; 128 AA.  
XX  
AC ABB70338;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 37806.

XX KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI: 2001-656860/75.

XX DR N-PSDB; ABL14441.

XX CC New isolated nucleic acid detection reagent for detecting 1000 or more PT genes from Drosophila and for elucidating cell signalling and cell-cell PT interactions -

XX PS Disclosure; SEQ ID NO 37806; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent CC capable of detecting 1000 or more genes from Drosophila. The invention is CC useful in developmental biology and in elucidating cell signalling and CC cell-cell interactions in higher eukaryotes for the development of CC insecticides, therapeutics and pharmaceutical drugs. The invention CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA CC sequences (ABL01840-ABL16175) and the encoded proteins CC (ABB57737-ABB72072).

XX CC The sequence data for this patent did not form part of the printed CC specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 128 AA;

Query Match 2.4%; Score 7; DB 22; Length 128;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 VITKNLI 68  
|||||

Db 34 VITKNLI 40

RESULT 35

AAW22483

ID AAW22483 standard; Protein; 158 AA.

AC AAW22483;

XX DT 10-MAR-1998 (first entry)

XX DE Phaffia derived glyceraldehyde-3-phosphate dehydrogenase PRcDNA10.

XX KW Glyceraldehyde-3-phosphate dehydrogenase; GAPDH; carotenoid;

XX KW synthesis; promoter; recombinant DNA; astaxanthin; ribosomal protein;

XX KW food colouring.

XX OS Phaffia rhodozyma.

XX PN WO9723633-A1.

XX PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

PD 03-JUL-1997.

XX 23-DEC-1996; 96WO-EP05887.

XX 11-APR-1996; 96EP-0200943.

XX 22-DEC-1995; 95EP-0203620.

XX (KONN) GIST-BROCADES BV.

XX PA (OOIJ) OOIJEN A J J.

XX PI Verdoes JC, Wery J;

XX DR WPI: 1997-351068/32.

XX DR N-PSDB; AAT72927.

XX CC Phaffia derived GAPDH and carotenoid synthesis genes and promoter PT fragment - used in the recombinant production of therapeutically PT useful proteins e.g. carotenoids for use in food colouring.

XX PS Claim 5; Page 75; 118pp; English.

XX CC The present sequence represents a Phaffia derived glyceraldehyde-3-phosphate dehydrogenase (GAPDH). The nucleic acid encoding this protein CC can be used in the novel recombinant DNA of the present invention. The CC recombinant DNA comprises a transcription promoter operably linked to a CC downstream sequence to be expressed where the transcription promoter CC comprises a region found upstream of the open reading frame (ORF) of a CC highly expressed Phaffia gene (preferably GAPDH, a ribosomal protein CC gene, an enzyme involved in the biosynthesis pathway). The recombinant CC DNA can be used to transform hosts, preferably Phaffia. These CC transformed hosts are then used in the recombinant production of GAPDH CC or an enzyme involved in carotenoid synthesis, preferably astaxanthin. CC They may also be used to produce a pharmaceutical product. Purified CC carotenoids can be used as colourants in food and/or feed, and also in CC cosmetics.

XX SQ Sequence 158 AA;

Query Match 2.4%; Score 7; DB 18; Length 158;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 YTFAGGT 120  
|||||

Db 148 YTFAGGT 154

RESULT 36

AAV60163

ID AAV60163 standard; Protein; 182 AA.

XX AC AAV60163;

XX DT 31-JAN-2000 (first entry)

XX DE Human endometrium tumour EST encoded protein 223.

XX KW Endometrium; human; tumour; cancer; anticancer; cytostatic; EST;

XX KW treatment; uterine; gene therapy; expressed sequence tag.

XX OS Homo sapiens.

XX PN DE19817948-A1.

XX PD 21-OCT-1999.

XX PF 17-APR-1998; 98DE-1017948.

XX PR 17-APR-1998; 98DE-1017948.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;



XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 249 AA;

Query Match 2.4%; Score 7; DB 22; Length 249;

Best Local Similarity 100.0%; Pred. No. 1.7e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 FSLNNLS 198

Db 150 FSLNNLS 156

RESULT 39

AAG58276

ID AAG58276 standard; Protein; 253 AA.

XX AAG58276;

AC AAG58276;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 75207.

KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

XX Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135623.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 21-JUL-1999; 99US-0145089.



```
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.4%; Score 7; DB 21; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 70 LKDVDS 76
Db 75 LKDVDS 81

RESULT 40
AAG61124
ID AAG61124 standard; Protein; 253 AA.
XX AAG61124;
AC AAG61124;
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 79247.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
```



PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 2.4%; Score 7; DB 21; Length 253;  
Best Local Similarity 100.0%; Pred.No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 LKKDVDS 76  
| | | | |  
Db 75 LKKDVDS 81

Search completed: July 8, 2003, 10:31:01  
Job time : 97 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 10:33:16 ; Search time 53 Seconds  
(without alignments)  
643.604 Million cell updates/sec

Title: US-10-062-624-40

Perfect score: 293

Sequence: 1 MNWKLKTIINTVLVCLLSL.....AVATINIGYFGGTCARLTF 293

Scoring table:

OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 445758 seqs, 116419773 residues

Word size : 0

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	293	100.0	293	9	US-10-062-051-40
2	293	100.0	293	9	US-10-062-920-40
3	293	100.0	293	12	US-10-062-624-40
4	168	57.3	293	9	US-10-314-639-44
5	168	57.3	293	12	US-10-059-964-44
6	25	8.5	293	9	US-10-284-986-13
7	25	8.5	293	10	US-09-846-808-13
8	25	8.5	297	9	US-10-314-639-14
9	25	8.5	297	12	US-10-059-964-14
10	24	8.2	296	9	US-10-314-639-56
11	24	8.2	296	12	US-10-059-964-56
12	22	7.5	298	9	US-10-284-986-11
13	22	7.5	298	9	US-10-314-639-52
14	22	7.5	298	10	US-09-846-808-11
15	22	7.5	298	12	US-10-059-964-52
16	17	5.8	300	9	US-10-284-986-12
17	17	5.8	300	9	US-10-314-639-50
18	17	5.8	300	10	US-09-846-808-12
19	17	5.8	300	12	US-10-059-964-50
20	16	5.5	291	9	US-10-284-986-10
21	16	5.5	291	9	US-10-314-639-18
22	16	5.5	291	10	US-09-846-808-10
23	16	5.5	291	12	US-10-059-964-18
24	15	5.1	241	9	US-10-314-639-62
25	15	5.1	241	12	US-10-059-964-62
26	15	5.1	275	9	US-10-284-986-8
27	15	5.1	275	9	US-10-314-639-28
28	15	5.1	275	10	US-09-846-808-8
29	15	5.1	275	12	US-10-059-964-28
30	13	4.4	281	9	US-10-314-639-58
31	13	4.4	281	12	US-10-059-964-58
32	11	3.8	279	9	US-10-284-986-6
33	11	3.8	279	9	US-10-314-639-24
34	11	3.8	279	10	US-09-846-808-6
35	11	3.8	279	12	US-10-059-964-24
36	11	3.8	283	9	US-10-284-986-7
37	11	3.8	283	9	US-10-314-639-26
38	11	3.8	283	10	US-09-846-808-7
39	11	3.8	283	12	US-10-059-964-26
40	11	3.8	295	9	US-10-284-986-5
41	11	3.8	295	9	US-10-314-639-22
42	11	3.8	295	10	US-09-846-808-5
43	11	3.8	295	12	US-10-059-964-22
44	11	3.8	299	9	US-10-314-639-46
45	11	3.8	299	12	US-10-059-964-46
46	10	3.4	285	9	US-10-284-986-9
47	10	3.4	285	9	US-10-314-639-30
48	10	3.4	285	10	US-09-846-808-9
49	10	3.4	285	12	US-10-059-964-30
50	10	3.4	293	9	US-10-314-639-54
51	10	3.4	293	12	US-10-059-964-54
52	9	3.1	279	9	US-10-314-639-60
53	9	3.1	279	12	US-10-059-964-60
54	8	2.7	131	9	US-10-314-639-20
55	8	2.7	131	12	US-10-059-964-20
56	8	2.7	229	9	US-10-286-264-20
57	8	2.7	272	9	US-10-284-986-4
58	8	2.7	272	10	US-09-846-808-4
59	8	2.7	280	9	US-10-062-051-42
60	8	2.7	280	9	US-10-062-920-42
61	8	2.7	280	9	US-10-314-639-48
62	8	2.7	280	12	US-10-062-624-42
63	8	2.7	280	12	US-10-059-964-48
64	7	2.4	20	9	US-10-098-093-53
65	7	2.4	101	10	US-09-864-761-45932
66	7	2.4	283	9	US-10-062-051-4
67	7	2.4	283	9	US-10-062-920-4
68	7	2.4	283	9	US-10-314-639-40
69	7	2.4	283	12	US-10-062-624-4
70	7	2.4	283	12	US-10-059-964-40
71	7	2.4	349	9	US-10-156-761-12266
72	7	2.4	505	9	US-10-076-622-478
73	7	2.4	505	9	US-10-076-622-485
74	7	2.4	505	9	US-09-551-621-478
75	7	2.4	505	10	US-09-604-287A-478
76	7	2.4	505	12	US-09-604-287A-485
77	7	2.4	505	12	US-10-007-805-478
78	7	2.4	505	12	US-10-007-805-485
79	7	2.4	518	9	US-09-738-626-5081
80	7	2.4	721	9	US-10-121-393-6
81	7	2.4	821	9	US-10-128-714-3444
82	7	2.4	860	9	US-10-128-714-8444
83	7	2.4	900	9	US-10-060-230-15
84	7	2.4	900	9	US-10-060-230-16
85	7	2.4	900	9	US-10-060-230-17
86	7	2.4	900	9	US-10-060-230-18
87	7	2.4	1025	10	US-09-912-020-283
88	6	2.0	16	10	US-09-842-164-7
89	6	2.0	28	10	US-09-929-818-105
90	6	2.0	38	10	US-09-864-761-34677
91	6	2.0	41	10	US-09-934-909-28
92	6	2.0	47	10	US-09-764-877-1625

Sequence 10, Appl  
Sequence 18, Appl  
Sequence 10, Appl  
Sequence 18, Appl  
Sequence 62, Appl  
Sequence 62, Appl  
Sequence 8, Appl  
Sequence 28, Appl  
Sequence 8, Appl  
Sequence 28, Appl  
Sequence 58, Appl  
Sequence 58, Appl  
Sequence 6, Appl  
Sequence 24, Appl  
Sequence 6, Appl  
Sequence 24, Appl  
Sequence 7, Appl  
Sequence 26, Appl  
Sequence 7, Appl  
Sequence 26, Appl  
Sequence 22, Appl  
Sequence 22, Appl  
Sequence 5, Appl  
Sequence 5, Appl  
Sequence 46, Appl  
Sequence 46, Appl  
Sequence 9, Appl  
Sequence 9, Appl  
Sequence 30, Appl  
Sequence 30, Appl  
Sequence 54, Appl  
Sequence 54, Appl  
Sequence 60, Appl  
Sequence 60, Appl  
Sequence 20, Appl  
Sequence 20, Appl  
Sequence 20, Appl  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 42, Appl  
Sequence 42, Appl  
Sequence 48, Appl  
Sequence 48, Appl  
Sequence 48, Appl  
Sequence 48, Appl  
Sequence 53, Appl  
Sequence 53, Appl  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 40, Appl  
Sequence 40, Appl  
Sequence 40, Appl  
Sequence 12266, A  
Sequence 478, App  
Sequence 478, App  
Sequence 478, App  
Sequence 485, App  
Sequence 485, App  
Sequence 485, App  
Sequence 485, App  
Sequence 581, App  
Sequence 6, Appl  
Sequence 3444, App  
Sequence 8444, App  
Sequence 15, Appl  
Sequence 15, Appl  
Sequence 16, Appl  
Sequence 17, Appl  
Sequence 18, Appl  
Sequence 283, App  
Sequence 7, Appl  
Sequence 105, App  
Sequence 34677, A  
Sequence 28, Appl  
Sequence 1625, App

93	6	2.0	48	9	US-10-083-357-952	Sequence 952, App	166	6	2.0	146	9	US-10-160-498-408	Sequence 408, App
94	6	2.0	48	10	US-09-864-761-36278	Sequence 36278, A	167	6	2.0	146	9	US-09-999-830A-221	Sequence 221, App
95	6	2.0	53	9	US-09-764-891-3131	Sequence 3131, App	168	6	2.0	146	9	US-10-121-041-408	Sequence 408, App
96	6	2.0	53	9	US-10-205-428-313	Sequence 313, App	169	6	2.0	146	9	US-10-121-043-408	Sequence 408, App
97	6	2.0	54	9	US-09-764-891-4474	Sequence 4474, App	170	6	2.0	146	9	US-10-121-047-408	Sequence 408, App
98	6	2.0	55	10	US-09-864-761-34547	Sequence 34547, A	171	6	2.0	146	9	US-10-123-215-408	Sequence 408, App
99	6	2.0	58	9	US-10-106-698-5593	Sequence 5593, App	172	6	2.0	146	9	US-10-123-902-408	Sequence 408, App
100	6	2.0	62	10	US-09-864-761-35362	Sequence 35362, A	173	6	2.0	146	9	US-10-123-908-408	Sequence 408, App
101	6	2.0	67	10	US-09-909-745-20	Sequence 20, App1	174	6	2.0	146	9	US-10-123-909-408	Sequence 408, App
102	6	2.0	81	10	US-09-925-302-502	Sequence 502, App	175	6	2.0	146	9	US-10-123-910-408	Sequence 408, App
103	6	2.0	95	9	US-10-194-975-91	Sequence 91, App1	176	6	2.0	146	9	US-10-124-813-408	Sequence 408, App
104	6	2.0	95	9	US-10-194-975-92	Sequence 92, App1	177	6	2.0	146	9	US-10-124-817-408	Sequence 408, App
105	6	2.0	96	9	US-10-153-382-34	Sequence 34, App1	178	6	2.0	146	9	US-10-124-824-408	Sequence 408, App
106	6	2.0	99	10	US-09-864-761-46578	Sequence 46578, A	179	6	2.0	146	9	US-10-125-922-408	Sequence 408, App
107	6	2.0	105	9	US-10-106-698-7530	Sequence 7530, App	180	6	2.0	146	9	US-10-125-924-408	Sequence 408, App
108	6	2.0	118	9	US-09-252-088-3	Sequence 3, App1	181	6	2.0	146	9	US-10-127-825A-408	Sequence 408, App
109	6	2.0	118	10	US-09-811-737-4	Sequence 4, App1	182	6	2.0	146	9	US-10-127-829A-408	Sequence 408, App
110	6	2.0	132	12	US-10-078-929-36	Sequence 36, App1	183	6	2.0	146	9	US-10-127-835A-408	Sequence 408, App
111	6	2.0	133	10	US-09-815-242-10372	Sequence 10372, A	184	6	2.0	146	9	US-10-127-839A-408	Sequence 408, App
112	6	2.0	133	10	US-09-815-242-13807	Sequence 13807, A	185	6	2.0	146	9	US-10-127-901A-408	Sequence 408, App
113	6	2.0	137	9	US-09-764-891-2740	Sequence 2740, App	186	6	2.0	146	9	US-10-128-693A-408	Sequence 408, App
114	6	2.0	140	9	US-09-791-932-93	Sequence 93, App1	187	6	2.0	146	9	US-10-131-813A-408	Sequence 408, App
115	6	2.0	144	9	US-10-101-464A-792	Sequence 792, App	188	6	2.0	146	9	US-10-131-818A-408	Sequence 408, App
116	6	2.0	146	9	US-09-978-295A-221	Sequence 221, App	189	6	2.0	146	9	US-10-131-823A-408	Sequence 408, App
117	6	2.0	146	9	US-09-978-697-221	Sequence 221, App	190	6	2.0	146	9	US-10-131-824A-408	Sequence 408, App
118	6	2.0	146	9	US-09-978-192A-221	Sequence 221, App	191	6	2.0	146	9	US-10-131-830A-408	Sequence 408, App
119	6	2.0	146	9	US-09-998-832A-221	Sequence 221, App	192	6	2.0	146	9	US-10-131-837A-408	Sequence 408, App
120	6	2.0	146	9	US-09-978-189-221	Sequence 221, App	193	6	2.0	146	9	US-10-137-872A-408	Sequence 408, App
121	6	2.0	146	9	US-10-028-072-408	Sequence 408, App	194	6	2.0	146	9	US-10-140-860-408	Sequence 408, App
122	6	2.0	146	9	US-10-121-049-408	Sequence 408, App	195	6	2.0	146	9	US-10-142-417-408	Sequence 408, App
123	6	2.0	146	9	US-10-123-904-408	Sequence 408, App	196	6	2.0	146	9	US-10-147-500-408	Sequence 408, App
124	6	2.0	146	9	US-10-140-470-408	Sequence 408, App	197	6	2.0	146	9	US-10-147-502-408	Sequence 408, App
125	6	2.0	146	9	US-10-175-746-408	Sequence 408, App	198	6	2.0	146	9	US-10-147-515-408	Sequence 408, App
126	6	2.0	146	9	US-10-176-921-408	Sequence 408, App	199	6	2.0	146	9	US-10-147-519-408	Sequence 408, App
127	6	2.0	146	9	US-10-176-921-408	Sequence 408, App	200	6	2.0	146	9	US-10-147-519-408	Sequence 408, App
128	6	2.0	146	9	US-10-137-865-408	Sequence 408, App	201	6	2.0	146	9	US-10-147-526-408	Sequence 408, App
129	6	2.0	146	9	US-10-140-474-408	Sequence 408, App	202	6	2.0	146	9	US-10-147-527-408	Sequence 408, App
130	6	2.0	146	9	US-10-142-431-408	Sequence 408, App	203	6	2.0	146	9	US-10-152-395-408	Sequence 408, App
131	6	2.0	146	9	US-10-143-114-408	Sequence 408, App	204	6	2.0	146	9	US-10-157-782-408	Sequence 408, App
132	6	2.0	146	9	US-10-140-003-408	Sequence 408, App	205	6	2.0	146	9	US-09-978-752A-221	Sequence 221, App
133	6	2.0	146	9	US-09-978-608A-221	Sequence 221, App	206	6	2.0	146	9	US-10-121-040-408	Sequence 408, App
134	6	2.0	146	9	US-10-142-419-408	Sequence 408, App	207	6	2.0	146	9	US-10-121-056-408	Sequence 408, App
135	6	2.0	146	9	US-09-978-191A-221	Sequence 221, App	208	6	2.0	146	9	US-10-121-061-408	Sequence 408, App
136	6	2.0	146	9	US-09-978-403A-221	Sequence 221, App	209	6	2.0	146	9	US-10-123-235-408	Sequence 408, App
137	6	2.0	146	9	US-09-978-564A-221	Sequence 221, App	210	6	2.0	146	9	US-10-124-818-408	Sequence 408, App
138	6	2.0	146	9	US-09-978-585A-221	Sequence 221, App	211	6	2.0	146	9	US-10-125-926A-408	Sequence 408, App
139	6	2.0	146	9	US-10-017-081A-221	Sequence 221, App	212	6	2.0	146	9	US-10-125-930A-408	Sequence 408, App
140	6	2.0	146	9	US-10-123-262-408	Sequence 408, App	213	6	2.0	146	9	US-10-127-831A-408	Sequence 408, App
141	6	2.0	146	9	US-10-142-423-408	Sequence 408, App	214	6	2.0	146	9	US-10-127-837A-408	Sequence 408, App
142	6	2.0	146	9	US-09-978-824-221	Sequence 221, App	215	6	2.0	146	9	US-10-127-838B-408	Sequence 408, App
143	6	2.0	146	9	US-09-981-915A-221	Sequence 221, App	216	6	2.0	146	9	US-10-127-842A-408	Sequence 408, App
144	6	2.0	146	9	US-09-999-833A-221	Sequence 221, App	217	6	2.0	146	9	US-10-127-843A-408	Sequence 408, App
145	6	2.0	146	9	US-10-121-050-408	Sequence 408, App	218	6	2.0	146	9	US-10-127-845A-408	Sequence 408, App
146	6	2.0	146	9	US-10-141-755-408	Sequence 408, App	219	6	2.0	146	9	US-10-127-846A-408	Sequence 408, App
147	6	2.0	146	9	US-10-167-749-221	Sequence 221, App	220	6	2.0	146	9	US-10-127-848A-408	Sequence 408, App
148	6	2.0	146	9	US-09-918-585A-221	Sequence 221, App	221	6	2.0	146	9	US-10-127-849A-408	Sequence 408, App
149	6	2.0	146	9	US-10-143-032-408	Sequence 408, App	222	6	2.0	146	9	US-10-127-850A-408	Sequence 408, App
150	6	2.0	146	9	US-09-978-423A-221	Sequence 221, App	223	6	2.0	146	9	US-10-127-851A-408	Sequence 408, App
151	6	2.0	146	9	US-10-013-921A-221	Sequence 221, App	224	6	2.0	146	9	US-10-128-684A-408	Sequence 408, App
152	6	2.0	146	9	US-10-123-108-408	Sequence 408, App	225	6	2.0	146	9	US-10-128-686A-408	Sequence 408, App
153	6	2.0	146	9	US-10-123-236-408	Sequence 408, App	226	6	2.0	146	9	US-10-128-690A-408	Sequence 408, App
154	6	2.0	146	9	US-10-123-261-408	Sequence 408, App	227	6	2.0	146	9	US-10-128-691A-408	Sequence 408, App
155	6	2.0	146	9	US-10-140-921-408	Sequence 408, App	228	6	2.0	146	9	US-10-131-819A-408	Sequence 408, App
156	6	2.0	146	9	US-10-140-928-408	Sequence 408, App	229	6	2.0	146	9	US-10-131-829A-408	Sequence 408, App
157	6	2.0	146	9	US-09-978-193A-221	Sequence 221, App	230	6	2.0	146	9	US-10-137-868-408	Sequence 408, App
158	6	2.0	146	9	US-10-013-929A-221	Sequence 221, App	231	6	2.0	146	9	US-10-137-868-408	Sequence 408, App
159	6	2.0	146	9	US-10-016-177A-221	Sequence 221, App	232	6	2.0	146	9	US-10-146-729-408	Sequence 408, App
160	6	2.0	146	9	US-10-121-045-408	Sequence 408, App	233	6	2.0	146	9	US-10-146-731-408	Sequence 408, App
161	6	2.0	146	9	US-10-123-292-408	Sequence 408, App	234	6	2.0	146	9	US-10-147-484-408	Sequence 408, App
162	6	2.0	146	9	US-10-123-903-408	Sequence 408, App	235	6	2.0	146	9	US-10-147-492-408	Sequence 408, App
163	6	2.0	146	9	US-10-124-819-408	Sequence 408, App	236	6	2.0	146	9	US-10-147-500-408	Sequence 408, App
164	6	2.0	146	9	US-10-124-822-408	Sequence 408, App	237	6	2.0	146	9	US-10-147-512-408	Sequence 408, App
165	6	2.0	146	9	US-10-140-925-408	Sequence 408, App	238	6	2.0	146	9	US-10-158-782-408	Sequence 408, App

239	6	2.0	146	9	US-10-175-735-408	Sequence 408, App	312	6	2.0	248	9	US-10-156-761-13383	Sequence 13383, A
240	6	2.0	146	9	US-10-123-905-408	Sequence 408, App	313	6	2.0	253	10	US-09-811-737-18	Sequence 18, App
241	6	2.0	146	9	US-10-123-907-408	Sequence 408, App	314	6	2.0	254	9	US-09-738-626-4681	Sequence 4681, App
242	6	2.0	146	9	US-10-124-815-408	Sequence 408, App	315	6	2.0	254	9	US-09-880-748-1469	Sequence 1469, App
243	6	2.0	146	9	US-10-125-921A-408	Sequence 408, App	316	6	2.0	255	10	US-09-811-737-15	Sequence 15, Appl
244	6	2.0	146	9	US-10-125-928A-408	Sequence 408, App	317	6	2.0	258	10	US-09-921-640-9	Sequence 9, Appl
245	6	2.0	146	9	US-10-127-821A-408	Sequence 408, App	318	6	2.0	262	10	US-09-811-737-19	Sequence 19, Appl
246	6	2.0	146	9	US-10-127-822A-408	Sequence 408, App	319	6	2.0	265	10	US-09-815-242-11851	Sequence 11851, A
247	6	2.0	146	9	US-10-127-826A-408	Sequence 408, App	320	6	2.0	267	10	US-09-747-835A-50	Sequence 50, Appl
248	6	2.0	146	9	US-10-127-826A-408	Sequence 408, App	321	6	2.0	267	10	US-09-747-835A-51	Sequence 51, Appl
249	6	2.0	146	9	US-10-127-827A-408	Sequence 408, App	322	6	2.0	270	12	US-10-078-929-38	Sequence 38, Appl
250	6	2.0	146	9	US-10-127-828A-408	Sequence 408, App	323	6	2.0	271	9	US-10-062-051-46	Sequence 46, Appl
251	6	2.0	146	9	US-10-127-830A-408	Sequence 408, App	324	6	2.0	271	9	US-10-062-920-46	Sequence 46, Appl
252	6	2.0	146	9	US-10-127-832A-408	Sequence 408, App	325	6	2.0	271	12	US-10-062-624-46	Sequence 46, Appl
253	6	2.0	146	9	US-10-127-833A-408	Sequence 408, App	326	6	2.0	277	9	US-09-895-913A-254	Sequence 254, App
254	6	2.0	146	9	US-10-127-834A-408	Sequence 408, App	327	6	2.0	277	9	US-10-156-761-11584	Sequence 11584, A
255	6	2.0	146	9	US-10-127-836A-408	Sequence 408, App	328	6	2.0	280	12	US-10-001-843-145	Sequence 145, App
256	6	2.0	146	9	US-10-127-841A-408	Sequence 408, App	329	6	2.0	283	9	US-10-062-051-10	Sequence 10, Appl
257	6	2.0	146	9	US-10-127-844A-408	Sequence 408, App	330	6	2.0	283	9	US-10-284-986-14	Sequence 14, Appl
258	6	2.0	146	9	US-10-128-687A-408	Sequence 408, App	331	6	2.0	283	9	US-10-062-920-10	Sequence 10, Appl
259	6	2.0	146	9	US-10-128-688A-408	Sequence 408, App	332	6	2.0	283	9	US-10-314-639-4	Sequence 4, Appl
260	6	2.0	146	9	US-10-128-689A-408	Sequence 408, App	333	6	2.0	283	10	US-09-846-808-14	Sequence 14, Appl
261	6	2.0	146	9	US-10-128-694A-408	Sequence 408, App	334	6	2.0	283	12	US-10-062-624-10	Sequence 10, Appl
262	6	2.0	146	9	US-10-131-825A-408	Sequence 408, App	335	6	2.0	283	12	US-10-059-964-4	Sequence 4, Appl
263	6	2.0	146	9	US-10-230-417-408	Sequence 408, App	336	6	2.0	285	9	US-10-284-986-3	Sequence 3, Appl
264	6	2.0	146	9	US-10-121-051-408	Sequence 408, App	337	6	2.0	288	10	US-09-846-808-3	Sequence 3, Appl
265	6	2.0	146	9	US-10-131-815A-408	Sequence 408, App	338	6	2.0	288	9	US-10-156-761-13799	Sequence 13799, A
266	6	2.0	146	9	US-10-131-817A-408	Sequence 408, App	339	6	2.0	294	10	US-09-815-242-5261	Sequence 5261, App
267	6	2.0	146	9	US-10-131-821A-408	Sequence 408, App	340	6	2.0	299	10	US-09-925-299-786	Sequence 786, App
268	6	2.0	146	9	US-10-131-822A-408	Sequence 408, App	341	6	2.0	301	9	US-10-106-698-4845	Sequence 4845, App
269	6	2.0	146	9	US-10-131-828A-408	Sequence 408, App	342	6	2.0	301	10	US-09-925-299-786	Sequence 786, App
270	6	2.0	146	9	US-10-131-835A-408	Sequence 408, App	343	6	2.0	303	9	US-10-156-761-9050	Sequence 9050, App
271	6	2.0	146	9	US-10-137-864A-408	Sequence 408, App	344	6	2.0	306	9	US-10-156-761-9786	Sequence 9786, App
272	6	2.0	146	9	US-10-137-869A-408	Sequence 408, App	345	6	2.0	308	9	US-09-805-919-1	Sequence 1, Appl
273	6	2.0	146	9	US-10-147-523-408	Sequence 408, App	346	6	2.0	310	10	US-09-815-242-10013	Sequence 10013, A
274	6	2.0	146	9	US-10-158-785-408	Sequence 408, App	347	6	2.0	316	10	US-09-799-777-49	Sequence 49, Appl
275	6	2.0	146	9	US-09-978-187B-221	Sequence 221, App	348	6	2.0	321	10	US-09-925-301-861	Sequence 861, App
276	6	2.0	146	9	US-10-121-042-408	Sequence 408, App	349	6	2.0	328	9	US-10-083-357-1324	Sequence 1324, App
277	6	2.0	146	9	US-10-123-912-408	Sequence 408, App	350	6	2.0	330	9	US-09-866-050A-144	Sequence 144, App
278	6	2.0	146	9	US-09-978-643A-221	Sequence 221, App	351	6	2.0	330	9	US-09-866-050A-178	Sequence 278, App
279	6	2.0	146	9	US-10-166-709A-221	Sequence 221, App	352	6	2.0	337	9	US-09-738-826-5808	Sequence 5808, App
280	6	2.0	146	9	US-10-192-007-408	Sequence 408, App	353	6	2.0	339	9	US-10-012-896-879	Sequence 879, App
281	6	2.0	146	9	US-10-194-359-408	Sequence 408, App	354	6	2.0	339	9	US-09-802-520-11	Sequence 11, Appl
282	6	2.0	146	9	US-10-127-847A-408	Sequence 408, App	355	6	2.0	339	9	US-09-895-793-879	Sequence 879, App
283	6	2.0	146	9	US-10-175-590-408	Sequence 408, App	356	6	2.0	339	9	US-09-895-814-879	Sequence 879, App
284	6	2.0	147	9	US-10-156-761-8595	Sequence 8595, App	357	6	2.0	339	9	US-10-011-095-2	Sequence 2, Appl
285	6	2.0	148	10	US-09-939-980-481	Sequence 481, App	358	6	2.0	339	9	US-10-010-867A-2	Sequence 2, Appl
286	6	2.0	151	10	US-09-789-561-103	Sequence 103, App	359	6	2.0	341	9	US-09-804-291-27	Sequence 27, Appl
287	6	2.0	154	10	US-09-867-550-950	Sequence 950, App	360	6	2.0	341	10	US-09-886-055-27	Sequence 27, Appl
288	6	2.0	155	9	US-10-153-382-35	Sequence 35, Appl	361	6	2.0	342	9	US-09-964-008-1	Sequence 1, Appl
289	6	2.0	156	9	US-09-862-540-59	Sequence 59, Appl	362	6	2.0	342	9	US-09-964-008-3	Sequence 3, Appl
290	6	2.0	156	10	US-09-939-980-389	Sequence 389, App	363	6	2.0	342	9	US-10-225-567A-643	Sequence 643, App
291	6	2.0	163	9	US-09-854-133-41	Sequence 41, Appl	364	6	2.0	342	9	US-10-333-844-2	Sequence 2, Appl
292	6	2.0	163	9	US-10-144-649A-41	Sequence 41, Appl	365	6	2.0	342	10	US-09-835-922-2	Sequence 2, Appl
293	6	2.0	163	10	US-09-738-973-41	Sequence 41, Appl	366	6	2.0	342	10	US-09-827-937A-2	Sequence 2, Appl
294	6	2.0	167	9	US-10-106-698-6225	Sequence 6225, App	367	6	2.0	342	10	US-09-780-576-2	Sequence 2, Appl
295	6	2.0	168	9	US-09-479-040-7	Sequence 7, Appl	368	6	2.0	344	9	US-10-232-686-9	Sequence 9, Appl
296	6	2.0	169	9	US-09-809-391-354	Sequence 354, App	369	6	2.0	344	9	US-09-820-843A-35	Sequence 35, Appl
297	6	2.0	171	10	US-09-789-561-175	Sequence 175, App	370	6	2.0	344	10	US-09-779-879A-9	Sequence 9, Appl
298	6	2.0	173	10	US-09-789-561-144	Sequence 144, App	371	6	2.0	344	10	US-09-779-880A-9	Sequence 9, Appl
299	6	2.0	179	10	US-09-909-745-22	Sequence 22, App	372	6	2.0	346	10	US-09-916-790-2	Sequence 2, Appl
300	6	2.0	187	9	US-09-764-891-5020	Sequence 5020, App	373	6	2.0	346	10	US-09-104-792-3	Sequence 3, Appl
301	6	2.0	187	9	US-10-205-428-432	Sequence 432, App	374	6	2.0	347	10	US-09-864-761-37014	Sequence 37014, A
302	6	2.0	189	10	US-09-925-300-1811	Sequence 1811, App	375	6	2.0	350	9	US-09-908-744-51	Sequence 51, Appl
303	6	2.0	195	10	US-09-925-300-1405	Sequence 1405, App	376	6	2.0	358	9	US-09-974-298-86	Sequence 86, Appl
304	6	2.0	199	10	US-09-921-640-6	Sequence 6, Appl	377	6	2.0	358	9	US-10-121-101B-1	Sequence 1, Appl
305	6	2.0	206	9	US-10-125-258-2	Sequence 2, Appl	378	6	2.0				
306	6	2.0	209	9	US-10-295-403-56	Sequence 56, Appl	379	6	2.0				
307	6	2.0	211	10	US-09-811-284-216	Sequence 216, App	380	6	2.0				
308	6	2.0	213	9	US-10-133-178-3	Sequence 3, Appl	381	6	2.0				
309	6	2.0	221	9	US-10-156-761-14864	Sequence 14864, A	382	6	2.0				
310	6	2.0	222	9	US-09-738-626-4464	Sequence 4464, App	383	6	2.0				
311	6	2.0	222	9	US-09-983-802-161	Sequence 161, App	384	6	2.0				

385	6	2.0	358	9	US-10-225-567A-418	Sequence 418, Appl	458	6	2.0	429	9	US-10-140-921-94	Sequence 94, Appl
386	6	2.0	358	10	US-09-826-508-10	Sequence 10, Appl	459	6	2.0	429	9	US-10-140-928-94	Sequence 94, Appl
387	6	2.0	358	10	US-09-861-451A-50	Sequence 50, Appl	460	6	2.0	429	9	US-10-121-045-94	Sequence 94, Appl
388	6	2.0	358	10	US-09-919-172-22	Sequence 22, Appl	461	6	2.0	429	9	US-10-123-292-94	Sequence 94, Appl
389	6	2.0	359	9	US-09-842-164-4	Sequence 4, Appli	462	6	2.0	429	9	US-10-123-903-94	Sequence 94, Appl
390	6	2.0	360	9	US-10-225-567A-460	Sequence 460, Appl	463	6	2.0	429	9	US-10-124-819-94	Sequence 94, Appl
391	6	2.0	360	10	US-09-131-827A-2	Sequence 2, Appli	464	6	2.0	429	9	US-10-124-822-94	Sequence 94, Appl
392	6	2.0	360	10	US-09-131-827A-20	Sequence 20, Appl	465	6	2.0	429	9	US-10-140-925-94	Sequence 94, Appl
393	6	2.0	360	10	US-09-938-719-7	Sequence 7, Appli	466	6	2.0	429	9	US-10-160-498-94	Sequence 94, Appl
394	6	2.0	360	10	US-09-939-226-7	Sequence 7, Appli	467	6	2.0	429	9	US-10-121-041-94	Sequence 94, Appl
395	6	2.0	360	10	US-09-938-703-7	Sequence 7, Appli	468	6	2.0	429	9	US-10-121-043-94	Sequence 94, Appl
396	6	2.0	361	9	US-09-738-626-5339	Sequence 5339, Ap	469	6	2.0	429	9	US-10-121-047-94	Sequence 94, Appl
397	6	2.0	363	10	US-09-815-242-12801	Sequence 12801, A	470	6	2.0	429	9	US-10-123-215-94	Sequence 94, Appl
398	6	2.0	371	9	US-10-099-278-64	Sequence 64, Appl	471	6	2.0	429	9	US-10-123-902-94	Sequence 94, Appl
399	6	2.0	372	9	US-10-156-761-13841	Sequence 13841, A	472	6	2.0	429	9	US-10-123-908-94	Sequence 94, Appl
400	6	2.0	376	9	US-10-156-761-13926	Sequence 13926, A	473	6	2.0	429	9	US-10-123-909-94	Sequence 94, Appl
401	6	2.0	379	9	US-10-156-761-11264	Sequence 11264, A	474	6	2.0	429	9	US-10-123-910-94	Sequence 94, Appl
402	6	2.0	384	9	US-10-292-525-12	Sequence 12, Appl	475	6	2.0	429	9	US-10-124-813-94	Sequence 94, Appl
403	6	2.0	384	9	US-10-225-567A-130	Sequence 130, Appl	476	6	2.0	429	9	US-10-124-817-94	Sequence 94, Appl
404	6	2.0	384	10	US-09-160-116-18	Sequence 18, Appl	477	6	2.0	429	9	US-10-124-824-94	Sequence 94, Appl
405	6	2.0	385	9	US-10-099-278-11	Sequence 11, Appl	478	6	2.0	429	9	US-10-125-922-94	Sequence 94, Appl
406	6	2.0	385	9	US-10-099-278-68	Sequence 68, Appl	479	6	2.0	429	9	US-10-125-924-94	Sequence 94, Appl
407	6	2.0	389	10	US-09-815-242-5336	Sequence 5336, Ap	480	6	2.0	429	9	US-10-127-825A-94	Sequence 94, Appl
408	6	2.0	391	9	US-10-099-278-78	Sequence 78, Appl	481	6	2.0	429	9	US-10-127-829A-94	Sequence 94, Appl
409	6	2.0	392	9	US-10-099-278-90	Sequence 90, Appl	482	6	2.0	429	9	US-10-127-835A-94	Sequence 94, Appl
410	6	2.0	395	9	US-10-219-220-182	Sequence 182, Appl	483	6	2.0	429	9	US-10-127-839A-94	Sequence 94, Appl
411	6	2.0	395	10	US-09-160-116-15	Sequence 15, Appl	484	6	2.0	429	9	US-10-127-901A-94	Sequence 94, Appl
412	6	2.0	398	10	US-09-815-242-12212	Sequence 12212, A	485	6	2.0	429	9	US-10-128-693A-94	Sequence 94, Appl
413	6	2.0	398	10	US-09-815-242-12912	Sequence 12912, A	486	6	2.0	429	9	US-10-131-813A-94	Sequence 94, Appl
414	6	2.0	398	10	US-09-815-242-13062	Sequence 13062, A	487	6	2.0	429	9	US-10-131-818A-94	Sequence 94, Appl
415	6	2.0	399	9	US-10-225-567A-58	Sequence 58, Appl	488	6	2.0	429	9	US-10-131-823A-94	Sequence 94, Appl
416	6	2.0	399	9	US-10-128-714-3462	Sequence 3462, Ap	489	6	2.0	429	9	US-10-131-824A-94	Sequence 94, Appl
417	6	2.0	399	10	US-09-160-116-12	Sequence 12, Appl	490	6	2.0	429	9	US-10-131-830A-94	Sequence 94, Appl
418	6	2.0	399	10	US-09-160-116-13	Sequence 13, Appl	491	6	2.0	429	9	US-10-131-837A-94	Sequence 94, Appl
419	6	2.0	399	10	US-09-160-116-14	Sequence 14, Appl	492	6	2.0	429	9	US-10-137-872A-94	Sequence 94, Appl
420	6	2.0	400	9	US-10-156-275-8	Sequence 8, Appli	493	6	2.0	429	9	US-10-140-860-94	Sequence 94, Appl
421	6	2.0	403	9	US-09-738-626-3904	Sequence 3904, Ap	494	6	2.0	429	9	US-10-142-417-94	Sequence 94, Appl
422	6	2.0	403	9	US-09-796-753-46	Sequence 46, Appl	495	6	2.0	429	9	US-10-147-500-94	Sequence 94, Appl
423	6	2.0	403	9	US-10-083-357-1289	Sequence 1289, Ap	496	6	2.0	429	9	US-10-147-502-94	Sequence 94, Appl
424	6	2.0	406	9	US-10-156-761-9851	Sequence 9851, Ap	497	6	2.0	429	9	US-10-147-515-94	Sequence 94, Appl
425	6	2.0	416	9	US-10-128-714-8462	Sequence 8462, Ap	498	6	2.0	429	9	US-10-147-517-94	Sequence 94, Appl
426	6	2.0	417	9	US-10-233-613-4	Sequence 4, Appli	499	6	2.0	429	9	US-10-147-519-94	Sequence 94, Appl
427	6	2.0	417	10	US-09-815-242-10165	Sequence 10165, A	500	6	2.0	429	9	US-10-147-526-94	Sequence 94, Appl
428	6	2.0	417	10	US-09-816-094-4	Sequence 4, Appli	501	6	2.0	429	9	US-10-147-527-94	Sequence 94, Appl
429	6	2.0	417	10	US-09-734-032-4	Sequence 4, Appli	502	6	2.0	429	9	US-10-152-395-94	Sequence 94, Appl
430	6	2.0	423	9	US-09-911-317-4	Sequence 2, Appli	503	6	2.0	429	9	US-10-157-782-94	Sequence 94, Appl
431	6	2.0	423	9	US-10-233-613-2	Sequence 2, Appli	504	6	2.0	429	9	US-10-121-040-94	Sequence 94, Appl
432	6	2.0	424	10	US-09-816-094-2	Sequence 2, Appli	505	6	2.0	429	9	US-10-121-056-94	Sequence 94, Appl
433	6	2.0	424	10	US-09-734-032-2	Sequence 2, Appli	506	6	2.0	429	9	US-10-121-061-94	Sequence 94, Appl
434	6	2.0	424	12	US-10-016-985-2	Sequence 2, Appli	507	6	2.0	429	9	US-10-123-335-94	Sequence 94, Appl
435	6	2.0	427	9	US-10-156-761-10902	Sequence 10902, A	508	6	2.0	429	9	US-10-124-818-94	Sequence 94, Appl
436	6	2.0	428	9	US-10-286-264-92	Sequence 92, Appl	509	6	2.0	429	9	US-10-125-926A-94	Sequence 94, Appl
437	6	2.0	429	9	US-10-028-072-94	Sequence 94, Appl	510	6	2.0	429	9	US-10-125-930A-94	Sequence 94, Appl
438	6	2.0	429	9	US-10-121-049-94	Sequence 94, Appl	511	6	2.0	429	9	US-10-127-831A-94	Sequence 94, Appl
439	6	2.0	429	9	US-10-123-904-94	Sequence 94, Appl	512	6	2.0	429	9	US-10-127-837A-94	Sequence 94, Appl
440	6	2.0	429	9	US-10-140-470-94	Sequence 94, Appl	513	6	2.0	429	9	US-10-127-838B-94	Sequence 94, Appl
441	6	2.0	429	9	US-10-175-746-94	Sequence 94, Appl	514	6	2.0	429	9	US-10-127-842A-94	Sequence 94, Appl
442	6	2.0	429	9	US-10-176-918-94	Sequence 94, Appl	515	6	2.0	429	9	US-10-127-843A-94	Sequence 94, Appl
443	6	2.0	429	9	US-10-176-921-94	Sequence 94, Appl	516	6	2.0	429	9	US-10-127-845A-94	Sequence 94, Appl
444	6	2.0	429	9	US-10-137-865-94	Sequence 94, Appl	517	6	2.0	429	9	US-10-127-846A-94	Sequence 94, Appl
445	6	2.0	429	9	US-10-140-474-94	Sequence 94, Appl	518	6	2.0	429	9	US-10-127-848A-94	Sequence 94, Appl
446	6	2.0	429	9	US-10-142-431-94	Sequence 94, Appl	519	6	2.0	429	9	US-10-127-849A-94	Sequence 94, Appl
447	6	2.0	429	9	US-10-143-114-94	Sequence 94, Appl	520	6	2.0	429	9	US-10-127-850A-94	Sequence 94, Appl
448	6	2.0	429	9	US-10-140-003-94	Sequence 94, Appl	521	6	2.0	429	9	US-10-127-851A-94	Sequence 94, Appl
449	6	2.0	429	9	US-10-142-419-94	Sequence 94, Appl	522	6	2.0	429	9	US-10-128-684A-94	Sequence 94, Appl
450	6	2.0	429	9	US-10-123-263-94	Sequence 94, Appl	523	6	2.0	429	9	US-10-128-686A-94	Sequence 94, Appl
451	6	2.0	429	9	US-10-142-423-94	Sequence 94, Appl	524	6	2.0	429	9	US-10-128-690A-94	Sequence 94, Appl
452	6	2.0	429	9	US-10-121-050-94	Sequence 94, Appl	525	6	2.0	429	9	US-10-128-691A-94	Sequence 94, Appl
453	6	2.0	429	9	US-10-141-753-94	Sequence 94, Appl	526	6	2.0	429	9	US-10-131-813A-94	Sequence 94, Appl
454	6	2.0	429	9	US-10-143-033-94	Sequence 94, Appl	527	6	2.0	429	9	US-10-131-829A-94	Sequence 94, Appl
455	6	2.0	429	9	US-10-123-108-94	Sequence 94, Appl	528	6	2.0	429	9	US-10-131-836A-94	Sequence 94, Appl
456	6	2.0	429	9	US-10-123-236-94	Sequence 94, Appl	529	6	2.0	429	9	US-10-137-868-94	Sequence 94, Appl
457	6	2.0	429	9	US-10-123-261-94	Sequence 94, Appl	530	6	2.0	429	9	US-10-146-729-94	Sequence 94, Appl



531	6	2.0	429	9	US-10-146-791-94	Sequence 94, Appl	604	510	9	US-10-029-180-94	Sequence 94, Appl
532	6	2.0	429	9	US-10-147-484-94	Sequence 94, Appl	605	511	10	US-09-815-242-13865	Sequence 13865, A
533	6	2.0	429	9	US-10-147-492-94	Sequence 94, Appl	606	511	10	US-10-097-340-212	Sequence 212, App
534	6	2.0	429	9	US-10-147-508-94	Sequence 94, Appl	607	515	9	US-10-171-311-156	Sequence 156, App
535	6	2.0	429	9	US-10-147-512-94	Sequence 94, Appl	608	515	9	US-10-263-819-2	Sequence 2, Appli
536	6	2.0	429	9	US-10-158-782-94	Sequence 94, Appl	609	520	9	US-10-205-823-433	Sequence 433, App
537	6	2.0	429	9	US-10-175-735-94	Sequence 94, Appl	610	520	9	US-10-081-051-60	Sequence 5050, Ap
538	6	2.0	429	9	US-10-123-905-94	Sequence 94, Appl	611	521	9	US-09-738-626-5902	Sequence 5902, Ap
539	6	2.0	429	9	US-10-123-907-94	Sequence 94, Appl	612	525	9	US-09-815-243-5902	Sequence 4, Appli
540	6	2.0	429	9	US-10-124-815-94	Sequence 94, Appl	613	525	12	US-10-011-436-4	Sequence 18, Appl
541	6	2.0	429	9	US-10-135-921A-94	Sequence 94, Appl	614	530	10	US-09-861-451A-18	Sequence 2, Appli
542	6	2.0	429	9	US-10-135-928A-94	Sequence 94, Appl	615	533	9	US-10-041-406-2	Sequence 27, Appl
543	6	2.0	429	9	US-10-127-821A-94	Sequence 94, Appl	616	533	9	US-10-162-435-27	Sequence 2, Appli
544	6	2.0	429	9	US-10-127-822A-94	Sequence 94, Appl	617	534	9	US-09-808-395-2	Sequence 9, Appli
545	6	2.0	429	9	US-10-127-833A-94	Sequence 94, Appl	618	534	9	US-09-821-616-9	Sequence 2, Appli
546	6	2.0	429	9	US-10-127-834A-94	Sequence 94, Appl	619	534	9	US-10-038-723-2	Sequence 32, Appl
547	6	2.0	429	9	US-10-127-826A-94	Sequence 94, Appl	620	540	10	US-09-729-454-32	Sequence 10275, A
548	6	2.0	429	9	US-10-127-827A-94	Sequence 94, Appl	621	540	10	US-09-815-243-11867	Sequence 2, Appl
549	6	2.0	429	9	US-10-127-828A-94	Sequence 94, Appl	622	545	10	US-09-815-242-10275	Sequence 11867, A
550	6	2.0	429	9	US-10-127-830A-94	Sequence 94, Appl	623	545	10	US-10-205-823-225	Sequence 10503, A
551	6	2.0	429	9	US-10-127-832A-94	Sequence 94, Appl	624	547	9	US-09-815-243-13458	Sequence 235, App
552	6	2.0	429	9	US-10-127-833A-94	Sequence 94, Appl	625	547	10	US-09-729-454-2	Sequence 13458, A
553	6	2.0	429	9	US-10-127-834A-94	Sequence 94, Appl	626	552	9	US-10-236-433-16	Sequence 2, Appl
554	6	2.0	429	9	US-10-127-836A-94	Sequence 94, Appl	627	552	9	US-10-236-433-16	Sequence 16, Appl
555	6	2.0	429	9	US-10-127-841A-94	Sequence 94, Appl	628	559	10	US-09-815-243-11867	Sequence 11867, A
556	6	2.0	429	9	US-10-127-844A-94	Sequence 94, Appl	629	559	10	US-09-815-242-10503	Sequence 10503, A
557	6	2.0	429	9	US-10-128-687A-94	Sequence 94, Appl	630	583	10	US-09-815-243-13458	Sequence 235, App
558	6	2.0	429	9	US-10-128-688A-94	Sequence 94, Appl	631	590	9	US-10-205-823-218	Sequence 218, App
559	6	2.0	429	9	US-10-128-689A-94	Sequence 94, Appl	632	590	9	US-10-325-567A-440	Sequence 440, App
560	6	2.0	429	9	US-10-128-694A-94	Sequence 94, Appl	633	590	9	US-10-156-761-10874	Sequence 10874, A
561	6	2.0	429	9	US-10-131-825A-94	Sequence 94, Appl	634	592	9	US-10-227-884-62	Sequence 62, Appl
562	6	2.0	429	9	US-10-230-417-94	Sequence 94, Appl	635	592	9	US-10-230-163-62	Sequence 62, Appl
563	6	2.0	429	9	US-10-121-051-94	Sequence 94, Appl	636	592	9	US-10-218-631-62	Sequence 62, Appl
564	6	2.0	429	9	US-10-131-815A-94	Sequence 94, Appl	637	592	9	US-10-230-338-62	Sequence 62, Appl
565	6	2.0	429	9	US-10-131-817A-94	Sequence 94, Appl	638	592	9	US-10-230-414-62	Sequence 62, Appl
566	6	2.0	429	9	US-10-131-821A-94	Sequence 94, Appl	639	592	9	US-10-216-159A-62	Sequence 62, Appl
567	6	2.0	429	9	US-10-131-822A-94	Sequence 94, Appl	640	592	9	US-10-218-849-62	Sequence 62, Appl
568	6	2.0	429	9	US-10-131-828A-94	Sequence 94, Appl	641	592	9	US-10-227-873-62	Sequence 62, Appl
569	6	2.0	429	9	US-10-131-835A-94	Sequence 94, Appl	642	592	9	US-10-227-883-62	Sequence 62, Appl
570	6	2.0	429	9	US-10-137-864A-94	Sequence 94, Appl	643	592	9	US-10-219-076-62	Sequence 62, Appl
571	6	2.0	429	9	US-10-137-869A-94	Sequence 94, Appl	644	592	9	US-10-230-434-62	Sequence 62, Appl
572	6	2.0	429	9	US-10-147-523-94	Sequence 94, Appl	645	592	9	US-10-219-003-62	Sequence 62, Appl
573	6	2.0	429	9	US-10-158-785-94	Sequence 94, Appl	646	592	9	US-10-219-075-62	Sequence 62, Appl
574	6	2.0	429	9	US-10-121-042-94	Sequence 94, Appl	647	592	9	US-10-219-464-62	Sequence 62, Appl
575	6	2.0	429	9	US-10-123-912-94	Sequence 94, Appl	648	592	9	US-10-219-466-62	Sequence 62, Appl
576	6	2.0	429	9	US-10-192-007-94	Sequence 94, Appl	649	592	9	US-10-219-479-62	Sequence 62, Appl
577	6	2.0	429	9	US-10-194-359-94	Sequence 94, Appl	650	592	9	US-10-219-481-62	Sequence 62, Appl
578	6	2.0	429	9	US-10-127-847A-94	Sequence 94, Appl	651	592	9	US-10-219-478-62	Sequence 62, Appl
579	6	2.0	433	9	US-09-738-626-6592	Sequence 6592, Ap	652	592	9	US-10-232-231-62	Sequence 62, Appl
580	6	2.0	433	9	US-10-102-806-463	Sequence 463, App	653	592	9	US-10-232-233-62	Sequence 62, Appl
581	6	2.0	443	10	US-09-801-368-174	Sequence 174, App	654	592	9	US-10-216-165-62	Sequence 62, Appl
582	6	2.0	444	9	US-10-081-816-59	Sequence 59, Appl	655	592	9	US-10-218-956-62	Sequence 62, Appl
583	6	2.0	444	9	US-10-156-761-7910	Sequence 7910, Ap	656	592	9	US-10-219-478-62	Sequence 62, Appl
584	6	2.0	454	10	US-09-815-242-10731	Sequence 10731, A	657	592	9	US-10-219-536-62	Sequence 62, Appl
585	6	2.0	457	9	US-10-156-761-13148	Sequence 13148, A	658	592	9	US-10-219-072-62	Sequence 62, Appl
586	6	2.0	460	9	US-09-738-626-6145	Sequence 6145, Ap	659	592	9	US-10-219-470-62	Sequence 62, Appl
587	6	2.0	462	9	US-10-099-278-82	Sequence 82, Appl	660	592	9	US-10-219-474-62	Sequence 62, Appl
588	6	2.0	465	10	US-09-815-242-5781	Sequence 5781, Ap	661	592	9	US-10-219-524-62	Sequence 62, Appl
589	6	2.0	468	10	US-09-815-242-12590	Sequence 12590, A	662	592	9	US-10-219-528-62	Sequence 62, Appl
590	6	2.0	469	9	US-10-093-766-59	Sequence 59, Appl	663	592	9	US-10-227-880-62	Sequence 62, Appl
591	6	2.0	469	9	US-10-097-340-171	Sequence 171, App	664	592	9	US-10-227-881-62	Sequence 62, Appl
592	6	2.0	469	9	US-10-097-340-173	Sequence 173, App	665	592	9	US-10-227-882-62	Sequence 62, Appl
593	6	2.0	470	10	US-09-881-752A-174	Sequence 174, App	666	592	9	US-10-230-436-62	Sequence 62, Appl
594	6	2.0	471	10	US-09-881-752A-18	Sequence 18, Appl	667	592	9	US-10-232-223-62	Sequence 62, Appl
595	6	2.0	480	10	US-09-776-529A-2	Sequence 2, Appli	668	592	9	US-10-232-225-62	Sequence 62, Appl
596	6	2.0	483	9	US-09-796-753-44	Sequence 44, Appl	669	592	9	US-10-232-227-62	Sequence 62, Appl
597	6	2.0	487	9	US-10-156-761-9670	Sequence 9670, Ap	670	592	9	US-10-232-229-62	Sequence 62, Appl
598	6	2.0	487	9	US-10-156-761-12237	Sequence 12237, A	671	592	9	US-10-232-234-62	Sequence 62, Appl
599	6	2.0	490	10	US-09-815-242-13162	Sequence 13162, A	672	592	9	US-10-233-205-62	Sequence 62, Appl
600	6	2.0	499	9	US-09-869-877-2	Sequence 2, Appli	673	592	9	US-10-216-160-62	Sequence 62, Appl
601	6	2.0	499	10	US-09-732-350-2	Sequence 2, Appli	674	592	9	US-10-216-162-62	Sequence 62, Appl
602	6	2.0	501	10	US-09-741-669-476	Sequence 476, App	675	592	9	US-10-216-164-62	Sequence 62, Appl
603	6	2.0	508	9	US-10-057-136-20	Sequence 20, Appl	676	592	9	US-10-216-167-62	Sequence 62, Appl
							677	592	9	US-10-216-168-62	Sequence 62, Appl

677	6	2.0	592	9	US-10-219-060-62	Sequence 62, Appl	750	6	2.0	802	9	US-10-237-535-78	Sequence 78, Appl
678	6	2.0	592	9	US-10-219-065-62	Sequence 62, Appl	751	6	2.0	802	9	US-10-238-183-78	Sequence 78, Appl
679	6	2.0	592	9	US-10-219-071-62	Sequence 62, Appl	752	6	2.0	802	9	US-10-238-283-78	Sequence 78, Appl
680	6	2.0	592	9	US-10-219-074-62	Sequence 62, Appl	753	6	2.0	802	9	US-10-238-370-78	Sequence 78, Appl
681	6	2.0	592	9	US-10-219-077-62	Sequence 62, Appl	754	6	2.0	802	9	US-10-245-055-78	Sequence 78, Appl
682	6	2.0	592	9	US-10-219-465-62	Sequence 62, Appl	755	6	2.0	802	9	US-10-245-147-78	Sequence 78, Appl
683	6	2.0	592	9	US-10-219-467-62	Sequence 62, Appl	756	6	2.0	802	9	US-10-245-730-78	Sequence 78, Appl
684	6	2.0	592	9	US-10-219-469-62	Sequence 62, Appl	757	6	2.0	802	9	US-10-245-739-78	Sequence 78, Appl
685	6	2.0	592	9	US-10-219-471-62	Sequence 62, Appl	758	6	2.0	802	9	US-10-246-210-78	Sequence 78, Appl
686	6	2.0	592	9	US-10-219-473-62	Sequence 62, Appl	759	6	2.0	802	9	US-10-239-196-78	Sequence 78, Appl
687	6	2.0	592	9	US-10-219-476-62	Sequence 62, Appl	760	6	2.0	802	9	US-10-243-024-78	Sequence 78, Appl
688	6	2.0	592	9	US-10-219-482-62	Sequence 62, Appl	761	6	2.0	802	9	US-10-243-409-78	Sequence 78, Appl
689	6	2.0	592	9	US-10-227-874-62	Sequence 62, Appl	762	6	2.0	802	9	US-10-245-033-78	Sequence 78, Appl
690	6	2.0	592	9	US-10-227-876-62	Sequence 62, Appl	763	6	2.0	802	9	US-10-245-621-78	Sequence 78, Appl
691	6	2.0	592	9	US-10-227-878-62	Sequence 62, Appl	764	6	2.0	802	9	US-10-245-890-78	Sequence 78, Appl
692	6	2.0	592	9	US-10-229-974-62	Sequence 62, Appl	765	6	2.0	802	9	US-10-243-095-78	Sequence 78, Appl
693	6	2.0	592	9	US-10-230-024-62	Sequence 62, Appl	766	6	2.0	802	9	US-10-245-185-78	Sequence 78, Appl
694	6	2.0	592	9	US-10-230-113-62	Sequence 62, Appl	767	6	2.0	802	9	US-10-245-427-78	Sequence 78, Appl
695	6	2.0	592	9	US-10-230-183-62	Sequence 62, Appl	768	6	2.0	802	9	US-10-245-473-78	Sequence 78, Appl
696	6	2.0	592	9	US-10-230-234-62	Sequence 62, Appl	769	6	2.0	802	9	US-10-245-770-78	Sequence 78, Appl
697	6	2.0	592	9	US-10-230-306-62	Sequence 62, Appl	770	6	2.0	802	9	US-10-245-877-78	Sequence 78, Appl
698	6	2.0	592	9	US-10-230-426-62	Sequence 62, Appl	771	6	2.0	802	9	US-10-246-976-78	Sequence 78, Appl
699	6	2.0	592	9	US-10-230-427-62	Sequence 62, Appl	772	6	2.0	802	9	US-10-243-320-78	Sequence 78, Appl
700	6	2.0	592	9	US-10-230-433-62	Sequence 62, Appl	773	6	2.0	802	9	US-10-242-743-78	Sequence 78, Appl
701	6	2.0	592	9	US-10-230-435-62	Sequence 62, Appl	774	6	2.0	802	9	US-10-242-845-78	Sequence 78, Appl
702	6	2.0	592	9	US-10-230-438-62	Sequence 62, Appl	775	6	2.0	802	9	US-10-237-636-78	Sequence 78, Appl
703	6	2.0	592	9	US-10-232-222-62	Sequence 62, Appl	776	6	2.0	802	9	US-10-238-325-78	Sequence 78, Appl
704	6	2.0	592	9	US-10-219-070-62	Sequence 62, Appl	777	6	2.0	802	9	US-10-238-346-78	Sequence 78, Appl
705	6	2.0	592	9	US-10-219-472-62	Sequence 62, Appl	778	6	2.0	802	9	US-10-238-411-78	Sequence 78, Appl
706	6	2.0	592	9	US-10-219-527-62	Sequence 62, Appl	779	6	2.0	802	9	US-10-243-124-78	Sequence 78, Appl
707	6	2.0	592	9	US-10-227-877-62	Sequence 62, Appl	780	6	2.0	802	9	US-10-243-425-78	Sequence 78, Appl
708	6	2.0	592	9	US-10-216-166-62	Sequence 62, Appl	781	6	2.0	802	9	US-10-243-446-78	Sequence 78, Appl
709	6	2.0	592	9	US-10-218-612-62	Sequence 62, Appl	782	6	2.0	802	9	US-10-245-874-78	Sequence 78, Appl
710	6	2.0	608	10	US-09-801-368-428	Sequence 428, Appl	783	6	2.0	802	9	US-10-242-653-78	Sequence 78, Appl
711	6	2.0	608	10	US-09-009-009-11	Sequence 11, Appl	784	6	2.0	802	9	US-10-243-167-78	Sequence 78, Appl
712	6	2.0	613	10	US-09-815-242-11755	Sequence 11755, A	785	6	2.0	802	9	US-10-243-388-78	Sequence 78, Appl
713	6	2.0	616	9	US-10-126-278-32	Sequence 32, Appl	786	6	2.0	802	9	US-10-244-947-78	Sequence 78, Appl
714	6	2.0	616	9	US-10-126-279-31	Sequence 31, Appl	787	6	2.0	802	9	US-10-244-968-78	Sequence 78, Appl
715	6	2.0	626	9	US-10-029-009-23	Sequence 23, Appl	788	6	2.0	802	9	US-10-244-990-78	Sequence 78, Appl
716	6	2.0	627	9	US-10-286-264-18	Sequence 18, Appl	789	6	2.0	802	9	US-10-245-079-78	Sequence 78, Appl
717	6	2.0	636	9	US-10-005-057A-9	Sequence 9, Appl	790	6	2.0	802	9	US-10-245-127-78	Sequence 78, Appl
718	6	2.0	640	9	US-10-038-723-13	Sequence 13, Appl	791	6	2.0	802	9	US-10-245-207-78	Sequence 78, Appl
719	6	2.0	645	10	US-09-919-172-41	Sequence 41, Appl	792	6	2.0	802	9	US-10-245-646-78	Sequence 78, Appl
720	6	2.0	645	10	US-09-925-300-1453	Sequence 1453, Ap	793	6	2.0	802	9	US-10-245-695-78	Sequence 78, Appl
721	6	2.0	655	9	US-10-101-464A-70	Sequence 70, Appl	794	6	2.0	802	9	US-10-245-699-78	Sequence 78, Appl
722	6	2.0	662	12	US-10-036-328A-8	Sequence 8, Appl	795	6	2.0	802	9	US-10-245-737-78	Sequence 78, Appl
723	6	2.0	670	9	US-10-128-714-3587	Sequence 3587, Ap	796	6	2.0	802	9	US-10-245-878-78	Sequence 78, Appl
724	6	2.0	672	9	US-10-156-761-8104	Sequence 8104, Ap	797	6	2.0	802	9	US-10-245-890-78	Sequence 78, Appl
725	6	2.0	676	10	US-09-771-161A-209	Sequence 209, Appl	798	6	2.0	802	9	US-10-245-899-78	Sequence 78, Appl
726	6	2.0	685	10	US-09-745-763-11	Sequence 11, Appl	799	6	2.0	802	9	US-10-245-900-78	Sequence 78, Appl
727	6	2.0	691	9	US-09-906-419-22	Sequence 22, Appl	800	6	2.0	802	9	US-10-247-058-78	Sequence 78, Appl
728	6	2.0	691	9	US-10-119-136-22	Sequence 22, Appl	801	6	2.0	802	9	US-10-237-471-78	Sequence 78, Appl
729	6	2.0	691	9	US-10-119-136-132	Sequence 132, Appl	802	6	2.0	802	9	US-10-238-261-78	Sequence 78, Appl
730	6	2.0	696	9	US-09-738-626-5259	Sequence 5259, Ap	803	6	2.0	802	9	US-10-238-324-78	Sequence 78, Appl
731	6	2.0	706	10	US-09-749-956-2	Sequence 2, Appl	804	6	2.0	802	9	US-10-241-860-78	Sequence 78, Appl
732	6	2.0	714	10	US-09-728-137-4	Sequence 4, Appl	805	6	2.0	802	9	US-10-242-172-78	Sequence 78, Appl
733	6	2.0	714	10	US-09-818-264-2	Sequence 2, Appl	806	6	2.0	802	9	US-10-242-652-78	Sequence 78, Appl
734	6	2.0	716	9	US-10-008-355-7	Sequence 7, Appl	807	6	2.0	802	9	US-10-242-990-78	Sequence 78, Appl
735	6	2.0	721	10	US-09-815-242-13768	Sequence 13768, A	808	6	2.0	802	9	US-10-243-023-78	Sequence 78, Appl
736	6	2.0	724	9	US-10-128-714-3025	Sequence 3025, Ap	809	6	2.0	802	9	US-10-243-103-78	Sequence 78, Appl
737	6	2.0	725	9	US-10-108-605-321	Sequence 321, Appl	810	6	2.0	802	9	US-10-243-276-78	Sequence 78, Appl
738	6	2.0	733	12	US-10-036-328A-4	Sequence 4, Appl	811	6	2.0	802	9	US-10-243-326-78	Sequence 78, Appl
739	6	2.0	734	10	US-09-815-242-10312	Sequence 10312, A	812	6	2.0	802	9	US-10-243-364-78	Sequence 78, Appl
740	6	2.0	749	9	US-10-156-761-15021	Sequence 15021, A	813	6	2.0	802	9	US-10-243-494-78	Sequence 78, Appl
741	6	2.0	753	10	US-09-884-889-6	Sequence 6, Appl	814	6	2.0	802	9	US-10-244-995-78	Sequence 78, Appl
742	6	2.0	764	9	US-10-045-624B-2	Sequence 2, Appl	815	6	2.0	802	9	US-10-245-230-78	Sequence 78, Appl
743	6	2.0	787	9	US-10-128-714-8587	Sequence 8587, Ap	816	6	2.0	802	9	US-10-245-253-78	Sequence 78, Appl
744	6	2.0	802	9	US-10-245-103-78	Sequence 78, Appl	817	6	2.0	802	9	US-10-245-454-78	Sequence 78, Appl
745	6	2.0	802	9	US-10-245-107-78	Sequence 78, Appl	818	6	2.0	802	9	US-10-245-479-78	Sequence 78, Appl
746	6	2.0	802	9	US-10-245-143-78	Sequence 78, Appl	819	6	2.0	802	9	US-10-245-499-78	Sequence 78, Appl
747	6	2.0	802	9	US-10-245-771-78	Sequence 78, Appl	820	6	2.0	802	9	US-10-245-772-78	Sequence 78, Appl
748	6	2.0	802	9	US-10-245-851-78	Sequence 78, Appl	821	6	2.0	802	9	US-10-245-811-78	Sequence 78, Appl
749	6	2.0	802	9	US-10-245-883-78	Sequence 78, Appl	822	6	2.0	802	9	US-10-245-812-78	Sequence 78, Appl

823	6	2.0	802	9	US-10-245-852-78	Sequence 78, Appl	896	6	2.0	1328	9	US-10-083-357-1306	Sequence 1306, Ap
824	6	2.0	802	9	US-10-245-875-78	Sequence 78, Appl	897	6	2.0	1328	9	US-10-083-357-1308	Sequence 1308, Ap
825	6	2.0	802	9	US-10-245-881-78	Sequence 78, Appl	898	6	2.0	1328	9	US-10-083-357-1321	Sequence 1321, Ap
826	6	2.0	802	9	US-10-245-911-78	Sequence 78, Appl	899	6	2.0	1328	9	US-10-083-357-1323	Sequence 1323, Ap
827	6	2.0	802	9	US-10-245-913-78	Sequence 78, Appl	900	6	2.0	1328	9	US-10-083-357-1328	Sequence 1328, Ap
828	6	2.0	802	9	US-10-246-080-78	Sequence 78, Appl	901	6	2.0	1328	9	US-10-083-357-1345	Sequence 1345, Ap
829	6	2.0	802	9	US-10-246-121-78	Sequence 78, Appl	902	6	2.0	1329	9	US-10-083-357-1346	Sequence 1346, Ap
830	6	2.0	802	9	US-10-246-305-78	Sequence 78, Appl	903	6	2.0	1465	9	US-10-083-357-1310	Sequence 1310, Ap
831	6	2.0	802	9	US-10-246-929-78	Sequence 78, Appl	904	6	2.0	1491	10	US-09-815-242-5568	Sequence 5568, Ap
832	6	2.0	802	9	US-10-247-036-78	Sequence 78, Appl	905	6	2.0	1502	10	US-09-815-242-12162	Sequence 12162, A
833	6	2.0	802	9	US-10-243-255-78	Sequence 78, Appl	906	6	2.0	1567	10	US-09-835-232-2	Sequence 2, Appli
834	6	2.0	802	9	US-10-245-810-78	Sequence 78, Appl	907	6	2.0	1577	9	US-10-219-834-150	Sequence 150, App
835	6	2.0	802	9	US-10-245-910-78	Sequence 78, Appl	908	6	2.0	1615	9	US-10-219-834-149	Sequence 149, App
836	6	2.0	802	9	US-10-246-098-78	Sequence 78, Appl	909	6	2.0	1617	9	US-10-090-453A-2	Sequence 2, Appli
837	6	2.0	803	9	US-10-201-481-19	Sequence 19, Appl	910	6	2.0	1617	9	US-10-005-338B-6	Sequence 6, Appli
838	6	2.0	827	10	US-09-815-242-11357	Sequence 11357, A	911	6	2.0	2165	9	US-09-923-070A-29	Sequence 29, Appl
839	6	2.0	842	9	US-09-291-417-91	Sequence 91, Appl	912	6	2.0	2165	9	US-09-923-070A-30	Sequence 30, Appl
840	6	2.0	846	9	US-10-303-683-22	Sequence 22, Appl	913	6	2.0	2165	9	US-09-161-122-29	Sequence 29, Appl
841	6	2.0	891	9	US-10-270-875-4	Sequence 4, Appli	914	6	2.0	2165	9	US-10-104-966-12	Sequence 12, Appl
842	6	2.0	891	9	US-10-270-878-4	Sequence 4, Appli	915	6	2.0	2227	9	US-10-135-988-2	Sequence 2, Appli
843	6	2.0	891	9	US-10-270-786-4	Sequence 4, Appli	916	6	2.0	2227	9	US-10-135-988-4	Sequence 4, Appli
844	6	2.0	891	9	US-10-270-710-4	Sequence 4, Appli	917	6	2.0	2227	9	US-10-135-988-6	Sequence 6, Appli
845	6	2.0	891	9	US-10-270-859-4	Sequence 4, Appli	918	6	2.0	2227	10	US-09-329-955-12	Sequence 12, Appl
846	6	2.0	892	9	US-10-270-875-2	Sequence 2, Appli	919	6	2.0	2862	10	US-09-742-659-5	Sequence 5, Appli
847	6	2.0	892	9	US-10-270-875-3	Sequence 3, Appli	920	6	2.0	2864	8	US-08-424-550B-394	Sequence 394, App
848	6	2.0	892	9	US-10-270-878-2	Sequence 2, Appli	921	6	2.0	2864	10	US-09-742-659-2	Sequence 2, Appli
849	6	2.0	892	9	US-10-270-878-3	Sequence 3, Appli	922	6	2.0	2865	10	US-09-742-659-6	Sequence 6, Appli
850	6	2.0	892	9	US-10-270-786-2	Sequence 2, Appli	923	6	2.0	4563	9	US-09-870-759-128	Sequence 128, App
851	6	2.0	892	9	US-10-270-786-3	Sequence 3, Appli	924	6	2.0	4563	9	US-09-802-640-32	Sequence 32, Appl
852	6	2.0	892	9	US-10-270-710-2	Sequence 2, Appli	925	6	2.0	4834	9	US-10-097-534-27	Sequence 27, Appl
853	6	2.0	892	9	US-10-270-710-3	Sequence 3, Appli	926	6	2.0	6146	9	US-10-156-761-10436	Sequence 10436, A
854	6	2.0	892	9	US-10-270-859-2	Sequence 2, Appli	927	6	2.0	6304	9	US-10-147-026-16	Sequence 16, Appl
855	6	2.0	892	9	US-10-270-859-3	Sequence 3, Appli	928	6	2.0	6	9	US-10-014-485A-139	Sequence 139, App
856	6	2.0	898	9	US-10-043-487-277	Sequence 277, App	929	5	1.7	7	10	US-09-322-261-431	Sequence 431, App
857	6	2.0	904	10	US-09-881-752A-158	Sequence 158, App	930	5	1.7	7	10	US-09-045-020-8	Sequence 8, Appli
858	6	2.0	916	9	US-09-738-626-6353	Sequence 6353, Ap	931	5	1.7	8	9	US-10-226-007-361	Sequence 361, App
859	6	2.0	920	10	US-09-815-242-11446	Sequence 11446, A	932	5	1.7	9	9	US-10-254-446A-143	Sequence 143, App
860	6	2.0	920	10	US-09-815-242-11604	Sequence 11604, A	933	5	1.7	10	9	US-09-860-395-13	Sequence 13, Appl
861	6	2.0	934	9	US-09-912-697-10	Sequence 10, Appl	934	5	1.7	10	9	US-09-860-395-14	Sequence 14, Appl
862	6	2.0	934	9	US-09-760-285-20	Sequence 20, Appl	935	5	1.7	10	9	US-09-876-904A-501	Sequence 501, App
863	6	2.0	934	10	US-09-788-657-19	Sequence 19, Appl	936	5	1.7	10	9	US-09-572-404B-2003	Sequence 2003, Ap
864	6	2.0	937	9	US-10-126-927-66	Sequence 66, Appl	937	5	1.7	10	9	US-09-572-404B-3072	Sequence 3072, Ap
865	6	2.0	937	9	US-10-126-931A-66	Sequence 66, Appl	938	5	1.7	10	9	US-09-572-404B-4176	Sequence 4176, Ap
866	6	2.0	943	10	US-09-815-242-10994	Sequence 10994, A	939	5	1.7	10	9	US-09-572-404B-4177	Sequence 4177, Ap
867	6	2.0	950	10	US-09-272-809-7	Sequence 7, Appli	940	5	1.7	10	9	US-10-062-848-41	Sequence 41, Appl
868	6	2.0	952	9	US-10-216-556-2	Sequence 2, Appli	941	5	1.7	10	9	US-10-226-007-362	Sequence 362, App
869	6	2.0	978	10	US-09-815-242-5381	Sequence 5381, Ap	942	5	1.7	10	9	US-10-226-007-373	Sequence 373, App
870	6	2.0	1032	10	US-09-815-242-12192	Sequence 12192, A	943	5	1.7	10	10	US-09-758-318-42	Sequence 42, App
871	6	2.0	1032	10	US-09-728-137-8	Sequence 8, Appli	944	5	1.7	10	10	US-09-860-276-13	Sequence 13, Appl
872	6	2.0	1044	10	US-09-728-137-2	Sequence 2, Appli	945	5	1.7	10	10	US-09-860-276-14	Sequence 14, Appl
873	6	2.0	1073	10	US-09-815-242-12361	Sequence 12361, A	946	5	1.7	10	10	US-09-982-172-12	Sequence 12, Appl
874	6	2.0	1077	9	US-10-121-911-1	Sequence 1, Appli	947	5	1.7	10	10	US-09-952-432A-12	Sequence 12, Appl
875	6	2.0	1085	9	US-10-274-990-4	Sequence 4, Appli	948	5	1.7	11	9	US-09-846-352-1	Sequence 1, Appli
876	6	2.0	1085	10	US-09-734-674-4	Sequence 4, Appli	949	5	1.7	11	9	US-10-116-252-44	Sequence 44, Appl
877	6	2.0	1088	10	US-09-920-804-2	Sequence 2, Appli	950	5	1.7	11	9	US-10-226-007-363	Sequence 363, App
878	6	2.0	1088	10	US-09-920-804-4	Sequence 4, Appli	951	5	1.7	11	9	US-10-226-007-374	Sequence 374, App
879	6	2.0	1122	9	US-10-072-094-8	Sequence 8, Appli	952	5	1.7	11	9	US-10-226-007-384	Sequence 384, App
880	6	2.0	1129	9	US-10-072-094-91	Sequence 91, Appl	953	5	1.7	11	9	US-10-226-007-384	Sequence 384, App
881	6	2.0	1129	9	US-10-274-990-2	Sequence 2, Appli	954	5	1.7	12	9	US-10-226-007-364	Sequence 281, App
882	6	2.0	1129	10	US-09-734-674-2	Sequence 2, Appli	955	5	1.7	12	9	US-10-226-007-364	Sequence 364, App
883	6	2.0	1138	12	US-10-036-328A-6	Sequence 6, Appli	956	5	1.7	12	9	US-10-226-007-375	Sequence 375, App
884	6	2.0	1139	9	US-10-128-714-8025	Sequence 8025, Ap	957	5	1.7	12	9	US-10-226-007-385	Sequence 385, App
885	6	2.0	1147	10	US-09-815-242-5468	Sequence 5468, Ap	958	5	1.7	12	9	US-10-226-007-394	Sequence 394, App
886	6	2.0	1210	12	US-10-036-328A-2	Sequence 2, Appli	959	5	1.7	13	9	US-10-226-007-365	Sequence 365, App
887	6	2.0	1211	9	US-10-156-761-9737	Sequence 9737, Ap	960	5	1.7	13	9	US-10-226-007-376	Sequence 376, App
888	6	2.0	1255	9	US-09-996-069-10	Sequence 10, Appl	961	5	1.7	13	9	US-10-226-007-386	Sequence 386, App
889	6	2.0	1255	9	US-10-171-311-158	Sequence 158, App	962	5	1.7	13	9	US-10-226-007-395	Sequence 395, App
890	6	2.0	1255	9	US-10-177-293-311	Sequence 311, App	963	5	1.7	13	9	US-10-226-007-403	Sequence 403, App
891	6	2.0	1322	9	US-10-083-357-1332	Sequence 1332, Ap	964	5	1.7	14	9	US-10-171-027-12	Sequence 12, Appl
892	6	2.0	1328	9	US-10-083-357-1287	Sequence 1287, Ap	965	5	1.7	14	9	US-10-171-027-13	Sequence 13, Appl
893	6	2.0	1328	9	US-10-083-357-1293	Sequence 1293, Ap	966	5	1.7	14	9	US-10-226-007-366	Sequence 366, App
894	6	2.0	1328	9	US-10-083-357-1295	Sequence 1295, Ap	967	5	1.7	14	9	US-10-226-007-377	Sequence 377, App
895	6	2.0	1328	9	US-10-083-357-1296	Sequence 1296, Ap	968	5	1.7	14	9	US-10-226-007-387	Sequence 387, App

```
969 5 1.7 14 9 US-10-226-007-396
970 5 1.7 14 9 US-10-226-007-404
971 5 1.7 14 9 US-10-226-007-411
972 5 1.7 14 9 US-10-106-804B-10
973 5 1.7 14 10 US-09-749-234A-7
974 5 1.7 14 10 US-09-848-664-3
975 5 1.7 15 9 US-09-948-227-9
976 5 1.7 15 9 US-09-907-969-492
977 5 1.7 15 9 US-09-907-969-497
978 5 1.7 15 9 US-10-023-282-582
979 5 1.7 15 9 US-10-226-007-367
980 5 1.7 15 9 US-10-226-007-378
981 5 1.7 15 9 US-10-226-007-388
982 5 1.7 15 9 US-10-226-007-397
983 5 1.7 15 9 US-10-226-007-405
984 5 1.7 15 9 US-10-226-007-412
985 5 1.7 15 9 US-10-226-007-418
986 5 1.7 15 9 US-10-198-053-492
987 5 1.7 15 9 US-10-198-053-497
988 5 1.7 16 9 US-09-968-561A-254
989 5 1.7 16 9 US-09-860-395-3
990 5 1.7 16 9 US-09-792-686A-18
991 5 1.7 16 9 US-10-226-007-368
992 5 1.7 16 9 US-10-226-007-379
993 5 1.7 16 9 US-10-226-007-389
994 5 1.7 16 9 US-10-226-007-398
995 5 1.7 16 9 US-10-226-007-406
996 5 1.7 16 9 US-10-226-007-413
997 5 1.7 16 9 US-10-226-007-419
998 5 1.7 16 9 US-10-226-007-424
999 5 1.7 16 9 US-10-106-698-6850
1000 5 1.7 16 10 US-09-192-854-143
```

## ALIGNMENTS

```
RESULT 1
US-10-062-051-40
; Sequence 40, Application US/10062051
; Publication No. US20030073095A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/10/062,051
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US/09/660,587
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 40
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis p28-1 protein
US-10-062-051-40

Query Match 100.0%; Score 293; DB 9; Length 293;
Best Local Similarity 100.0%; Pred. No. 3.1e-279;
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKLKFTIINTVLVCLLSLPNISSKAINNAKKYGLYISGQYKPSVSVFSNFSVKET 60
DB 1 MNKLKFTIINTVLVCLLSLPNISSKAINNAKKYGLYISGQYKPSVSVFSNFSVKET 60
QY 61 NVITKNLIALKKVDSDIETKTDAISVGISNPSNFTIPYTAVFQDQNSVNFNGTIGYFAEQT 120
|||||
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/10/062,051
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US/09/660,587
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 40
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis p28-1 protein
US-10-062-051-40

Query Match 100.0%; Score 293; DB 9; Length 293;
Best Local Similarity 100.0%; Pred. No. 3.1e-279;
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKLKFTIINTVLVCLLSLPNISSKAINNAKKYGLYISGQYKPSVSVFSNFSVKET 60
DB 1 MNKLKFTIINTVLVCLLSLPNISSKAINNAKKYGLYISGQYKPSVSVFSNFSVKET 60
QY 61 NVITKNLIALKKVDSDIETKTDAISVGISNPSNFTIPYTAVFQDQNSVNFNGTIGYFAEQT 120
|||||
```

```
Db 61 NVITKNLIALKKVDSDIETKTDAISVGISNPSNFTIPYTAVFQDQNSVNFNGTIGYFAEQT 120
QY 121 RVEIEGSEYEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVKNSIFHTVMRNDGLS 180
121 RVEIEGSEYEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVKNSIFHTVMRNDGLS 180
121 RVEIEGSEYEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVKNSIFHTVMRNDGLS 180
181 IISVIVNVVDFSLNNLSISPIYICGGAGVDAIEFPDVLHIKFAYSQSLGIAYSLPSNISL 240
181 IISVIVNVVDFSLNNLSISPIYICGGAGVDAIEFPDVLHIKFAYSQSLGIAYSLPSNISL 240
241 FASLYYHKVMGNQPKNLNVQHVLAELASIPKITSAVATLNIGYFGGEIGARLTF 293
|||||
241 FASLYYHKVMGNQPKNLNVQHVLAELASIPKITSAVATLNIGYFGGEIGARLTF 293

RESULT 2
US-10-062-920-40
; Sequence 40, Application US/10062920
; Publication No. US20030096250A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/10/062,920
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US/09/660,587
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 40
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis p28-1 protein
US-10-062-920-40

Query Match 100.0%; Score 293; DB 9; Length 293;
Best Local Similarity 100.0%; Pred. No. 3.1e-279;
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKLKFTIINTVLVCLLSLPNISSKAINNAKKYGLYISGQYKPSVSVFSNFSVKET 60
DB 1 MNKLKFTIINTVLVCLLSLPNISSKAINNAKKYGLYISGQYKPSVSVFSNFSVKET 60
QY 61 NVITKNLIALKKVDSDIETKTDAISVGISNPSNFTIPYTAVFQDQNSVNFNGTIGYFAEQT 120
|||||
61 NVITKNLIALKKVDSDIETKTDAISVGISNPSNFTIPYTAVFQDQNSVNFNGTIGYFAEQT 120
|||||
QY 121 RVEIEGSEYEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVKNSIFHTVMRNDGLS 180
121 RVEIEGSEYEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVKNSIFHTVMRNDGLS 180
181 IISVIVNVVDFSLNNLSISPIYICGGAGVDAIEFPDVLHIKFAYSQSLGIAYSLPSNISL 240
181 IISVIVNVVDFSLNNLSISPIYICGGAGVDAIEFPDVLHIKFAYSQSLGIAYSLPSNISL 240
241 FASLYYHKVMGNQPKNLNVQHVLAELASIPKITSAVATLNIGYFGGEIGARLTF 293
|||||
241 FASLYYHKVMGNQPKNLNVQHVLAELASIPKITSAVATLNIGYFGGEIGARLTF 293
```

```
RESULT 3
US-10-062-624-40
; Sequence 40, Application US/10062624
; Patent No. US20020115840A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
```

```
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-Kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2/D1
; CURRENT APPLICATION NUMBER: US/10/062,624
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 09/660,587
; PRIOR FILING DATE: 2000-09-12
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 40
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis p28-1 protein
US-10-062-624-40

Query Match      100.0%; Score 293; DB 12; Length 293;
Best Local Similarity 100.0%; Pred. No. 3.1e-279; Indels 0; Gaps 0;
Matches 293; Conservative 0; Mismatches 0;

Qy 1 MNNKLKFTIINTVLVCLSLPNISSSKAINNNAKKYGLYISGOYKPSVSFNSFKET 60
Db 1 MNNKLKFTIINTVLVCLSLPNISSSKAINNNAKKYGLYISGOYKPSVSFNSFKET 60

Qy 61 NVITKNLALKQVDSTETKTDSVIGSNPSNFTIPYTAVFQDNSVNFNGTIGYTFABGT 120
Db 61 NVITKNLALKQVDSTETKTDSVIGSNPSNFTIPYTAVFQDNSVNFNGTIGYTFABGT 120

Qy 121 RVEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNS 180
Db 121 RVEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNS 180

Qy 181 IISVIVNVCYDFSLNNLSISPIYICGAGVDAIEFFDVLHIFAYQSKLGIAYSPLSNISL 240
Db 181 IISVIVNVCYDFSLNNLSISPIYICGAGVDAIEFFDVLHIFAYQSKLGIAYSPLSNISL 240

Qy 241 FASLYYHKVMGNQFNLMVQVHAEIASIPKITSAVATLNIGYFGGEIGARLTF 293
Db 241 FASLYYHKVMGNQFNLMVQVHAEIASIPKITSAVATLNIGYFGGEIGARLTF 293

RESULT 4
US-10-314-639-44
; Sequence 44, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/314,701
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-10-314-639-44

Query Match      57.3%; Score 168; DB 9; Length 293;
Best Local Similarity 100.0%; Pred. No. 1.3e-156; Indels 0; Gaps 0;
Matches 168; Conservative 0; Mismatches 0;

Qy 1 MNNKLKFTIINTVLVCLSLPNISSSKAINNNAKKYGLYISGOYKPSVSFNSFKET 60
Db 1 MNNKLKFTIINTVLVCLSLPNISSSKAINNNAKKYGLYISGOYKPSVSFNSFKET 60

Qy 61 NVITKNLALKQVDSTETKTDSVIGSNPSNFTIPYTAVFQDNSVNFNGTIGYTFABGT 120
Db 61 NVITKNLALKQVDSTETKTDSVIGSNPSNFTIPYTAVFQDNSVNFNGTIGYTFABGT 120

Qy 121 RVEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNS 180
Db 121 RVEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNS 180

Qy 181 IISVIVNVCYDFSLNNLSISPIYICGAGVDAIEFFDVLHIFAYQSKLGIAYSPLSNISL 240
Db 181 IISVIVNVCYDFSLNNLSISPIYICGAGVDAIEFFDVLHIFAYQSKLGIAYSPLSNISL 240

Qy 241 FASLYYHKVMGNQFNLMVQVHAEIASIPKITSAVATLNIGYFGGEIGARLTF 293
Db 241 FASLYYHKVMGNQFNLMVQVHAEIASIPKITSAVATLNIGYFGGEIGARLTF 293

RESULT 5
US-10-059-964-44
; Sequence 44, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; CURRENT FILING DATE: 2002-01-28
; EARLIER APPLICATION NUMBER: 09/314,701
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-10-059-964-44

Query Match      57.3%; Score 168; DB 12; Length 293;
Best Local Similarity 100.0%; Pred. No. 1.3e-156; Indels 0; Gaps 0;
Matches 168; Conservative 0; Mismatches 0;

Qy 1 MNNKLKFTIINTVLVCLSLPNISSSKAINNNAKKYGLYISGOYKPSVSFNSFKET 60
Db 1 MNNKLKFTIINTVLVCLSLPNISSSKAINNNAKKYGLYISGOYKPSVSFNSFKET 60

Qy 61 NVITKNLALKQVDSTETKTDSVIGSNPSNFTIPYTAVFQDNSVNFNGTIGYTFABGT 120
Db 61 NVITKNLALKQVDSTETKTDSVIGSNPSNFTIPYTAVFQDNSVNFNGTIGYTFABGT 120

Qy 121 RVEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNS 180
Db 121 RVEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNS 180

Qy 181 IISVIVNVCYDFSLNNLSISPIYICGAGVDAIEFFDVLHIFAYQSKLGIAYSPLSNISL 240
Db 181 IISVIVNVCYDFSLNNLSISPIYICGAGVDAIEFFDVLHIFAYQSKLGIAYSPLSNISL 240

Qy 241 FASLYYHKVMGNQFNLMVQVHAEIASIPKITSAVATLNIGYFGGEIGARLTF 293
Db 241 FASLYYHKVMGNQFNLMVQVHAEIASIPKITSAVATLNIGYFGGEIGARLTF 293

RESULT 6
US-10-284-986-13
; Sequence 13, Application US/10284986
; Publication No. US20030091588A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xu-Jie
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane
; FILE REFERENCE: D6311D1
; CURRENT APPLICATION NUMBER: US/10/284,986
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: 09/846,808
; PRIOR FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 13
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: P28-13 Outer Membrane Protein of
; OTHER INFORMATION: Ehrlichia chaffeensis
US-10-284-986-13

Query Match      8.5%; Score 25; DB 9; Length 293;
Best Local Similarity 100.0%; Pred. No. 2.5e-16;
```

```
Db 61 NVITKNLALKQVDSTETKTDSVIGSNPSNFTIPYTAVFQDNSVNFNGTIGYTFABGT 120
Qy 121 RVEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNS 168
Db 121 RVEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNS 168

RESULT 5
US-10-059-964-44
; Sequence 44, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; CURRENT FILING DATE: 2002-01-28
; EARLIER APPLICATION NUMBER: 09/314,701
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-10-059-964-44

Query Match      57.3%; Score 168; DB 12; Length 293;
Best Local Similarity 100.0%; Pred. No. 1.3e-156; Indels 0; Gaps 0;
Matches 168; Conservative 0; Mismatches 0;

Qy 1 MNNKLKFTIINTVLVCLSLPNISSSKAINNNAKKYGLYISGOYKPSVSFNSFKET 60
Db 1 MNNKLKFTIINTVLVCLSLPNISSSKAINNNAKKYGLYISGOYKPSVSFNSFKET 60

Qy 61 NVITKNLALKQVDSTETKTDSVIGSNPSNFTIPYTAVFQDNSVNFNGTIGYTFABGT 120
Db 61 NVITKNLALKQVDSTETKTDSVIGSNPSNFTIPYTAVFQDNSVNFNGTIGYTFABGT 120

Qy 121 RVEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNS 168
Db 121 RVEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNS 168

RESULT 6
US-10-284-986-13
; Sequence 13, Application US/10284986
; Publication No. US20030091588A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xu-Jie
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane
; FILE REFERENCE: D6311D1
; CURRENT APPLICATION NUMBER: US/10/284,986
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: 09/846,808
; PRIOR FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 13
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: P28-13 Outer Membrane Protein of
; OTHER INFORMATION: Ehrlichia chaffeensis
US-10-284-986-13

Query Match      8.5%; Score 25; DB 9; Length 293;
Best Local Similarity 100.0%; Pred. No. 2.5e-16;
```

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GLYISGQYKPSVSVFNSFSVKETNV 62  
|||||  
Db 38 GLYISGQYKPSVSVFNSFSVKETNV 62

## RESULT 7

US-09-846-808-13  
; Sequence 13, Application US/09946808  
; Patent No. US20020064531A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: Yu, Xu-Jie  
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane  
; FILE OF INVENTION: Protein Multigene Family  
; FILE REFERENCE: D6311  
; CURRENT APPLICATION NUMBER: US/09/846,808  
; CURRENT FILING DATE: 2001-05-01  
; PRIOR APPLICATION NUMBER: 60/201,035  
; PRIOR FILING DATE: 2000-05-01  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 13  
; LENGTH: 293  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: P28-13 Outer Membrane Protein of  
; OTHER INFORMATION: Ehrlichia chaffeensis  
US-09-846-808-13

Query Match 8.5%; Score 25; DB 10; Length 293;  
Best Local Similarity 100.0%; Pred. No. 2.5e-16;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GLYISGQYKPSVSVFNSFSVKETNV 62  
|||||  
Db 38 GLYISGQYKPSVSVFNSFSVKETNV 62

## RESULT 8

US-10-314-639-14  
; Sequence 14, Application US/10314639  
; Publication No. US20030103991A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20030103991A1io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; FILE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/314,639  
; CURRENT FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US/09/314,701  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 297  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
US-10-314-639-14

Query Match 8.5%; Score 25; DB 9; Length 297;  
Best Local Similarity 100.0%; Pred. No. 2.5e-16;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GLYISGQYKPSVSVFNSFSVKETNV 62  
|||||  
Db 42 GLYISGQYKPSVSVFNSFSVKETNV 66

## RESULT 9

US-10-059-964-14

; Sequence 14, Application US/10059964  
; Patent No. US20020120115A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20020120115A1io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; FILE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/059,964  
; CURRENT FILING DATE: 2002-01-28  
; PRIOR APPLICATION NUMBER: 09/314,701  
; EARLIER FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 297  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
US-10-059-964-14

Query Match 8.5%; Score 25; DB 12; Length 297;  
Best Local Similarity 100.0%; Pred. No. 2.5e-16;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GLYISGQYKPSVSVFNSFSVKETNV 62  
|||||  
Db 42 GLYISGQYKPSVSVFNSFSVKETNV 66

## RESULT 10

US-10-314-639-56  
; Sequence 56, Application US/10314639  
; Publication No. US20030103991A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20030103991A1io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; FILE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/314,639  
; CURRENT FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US/09/314,701  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 56  
; LENGTH: 296  
; TYPE: PRT  
; ORGANISM: p30-7  
US-10-314-639-56

Query Match 8.2%; Score 24; DB 9; Length 296;  
Best Local Similarity 100.0%; Pred. No. 2.4e-15;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GLYISGQYKPSVSVFNSFSVKETN 61  
|||||  
Db 39 GLYISGQYKPSVSVFNSFSVKETN 62

## RESULT 11

US-10-059-964-56  
; Sequence 56, Application US/10059964  
; Patent No. US20020120115A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20020120115A1io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; FILE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/059,964  
; CURRENT FILING DATE: 2002-01-28  
; EARLIER APPLICATION NUMBER: 09/314,701

Query Match  
Best Local Similarity 100.0%; Score 24; DB 12; Length 296;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
; EARLIER FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 56  
; LENGTH: 296  
; TYPE: PRT  
; ORGANISM: P30-7  
US-10-059-964-56

Query Match  
Best Local Similarity 100.0%; Score 24; DB 12; Length 296;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 GLYISGQYKPSVSVFNSFVSKETN 61  
Db 39 GLYISGQYKPSVSVFNSFVSKETN 62

## RESULT 12

US-10-284-986-11  
; Sequence 11, Application US/10284986  
; Publication No. US20030091588A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: Yu, Xu-Jie  
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane  
; FILE REFERENCE: D6311D1  
; CURRENT APPLICATION NUMBER: US/10/284,986  
; CURRENT FILING DATE: 2002-10-31  
; PRIOR APPLICATION NUMBER: 09/846,808  
; PRIOR FILING DATE: 2001-05-01  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 11  
; LENGTH: 298  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: P28-11 Outer Membrane Protein of  
; OTHER INFORMATION: Ehrlichia chaffeensis  
US-10-284-986-11

Query Match  
Best Local Similarity 100.0%; Score 22; DB 9; Length 298;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 40 YISGQYKPSVSVFNSFVSKETN 61  
Db 39 YISGQYKPSVSVFNSFVSKETN 60

## RESULT 13

US-10-314-639-52  
; Sequence 52, Application US/10314639  
; Publication No. US20030103991A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20030103991A1  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/314,639  
; CURRENT FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US/09/314,701  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 52  
; LENGTH: 298  
; TYPE: PRT  
; ORGANISM: OMP-1H  
US-10-314-639-52

Query Match  
Best Local Similarity 100.0%; Score 22; DB 9; Length 298;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
; EARLIER FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 56  
; LENGTH: 296  
; TYPE: PRT  
; ORGANISM: P30-7  
US-10-059-964-56

Qy 40 YISGQYKPSVSVFNSFVSKETN 61  
Db 39 YISGQYKPSVSVFNSFVSKETN 60

## RESULT 14

US-09-846-808-11  
; Sequence 11, Application US/09846808  
; Patent No. US20020064531A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: Yu, Xu-Jie  
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane  
; FILE REFERENCE: D6311  
; CURRENT APPLICATION NUMBER: US/09/846,808  
; CURRENT FILING DATE: 2001-05-01  
; PRIOR APPLICATION NUMBER: 60/201,035  
; PRIOR FILING DATE: 2000-05-01  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 11  
; LENGTH: 298  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: P28-11 Outer Membrane Protein of  
; OTHER INFORMATION: Ehrlichia chaffeensis  
US-09-846-808-11

Query Match  
Best Local Similarity 100.0%; Score 22; DB 10; Length 298;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 40 YISGQYKPSVSVFNSFVSKETN 61  
Db 39 YISGQYKPSVSVFNSFVSKETN 60

## RESULT 15

US-10-059-964-52  
; Sequence 52, Application US/10059964  
; Patent No. US20020120115A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20020120115A1  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/059,964  
; CURRENT FILING DATE: 2002-01-28  
; EARLIER APPLICATION NUMBER: 09/314,701  
; EARLIER FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 52  
; LENGTH: 298  
; TYPE: PRT  
; ORGANISM: OMP-1H  
US-10-059-964-52

Query Match  
Best Local Similarity 100.0%; Score 22; DB 12; Length 298;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 40 YISGQYKPSVSVFNSFVSKETN 61  
Db 39 YISGQYKPSVSVFNSFVSKETN 60

```
RESULT 16
US-10-284-986-12
; Sequence 12, Application US/10284986
; Publication No. US20030091588A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xu-Jie
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane
; FILE REFERENCE: D6311D1
; CURRENT APPLICATION NUMBER: US/10/284,986
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: 09/846,808
; PRIOR FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 12
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: P28-12 Outer Membrane Protein of
; OTHER INFORMATION: Ehrlichia chaffeensis
US-10-284-986-12

Query Match      5.8%; Score 17; DB 9; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      249 VMGNQFKNLNVQHVHVEL 265
Db      256 VMGNQFKNLNVQHVHVEL 272
|||||

RESULT 17
US-10-314-639-50
; Sequence 50, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20030103991A1
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/314,701
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn.Ver. 2.0
; SEQ ID NO 50
; LENGTH: 300
; TYPE: PRT
; ORGANISM: OMP-12
US-10-314-639-50

Query Match      5.8%; Score 17; DB 9; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      249 VMGNQFKNLNVQHVHVEL 265
Db      256 VMGNQFKNLNVQHVHVEL 272
|||||

RESULT 18
US-09-846-808-12
; Sequence 12, Application US/09846808
; Patent No. US20020064531A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xu-Jie
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane
; FILE REFERENCE: D6311D1
; CURRENT APPLICATION NUMBER: US/10/284,986
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: 09/846,808
; PRIOR FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 12
; LENGTH: 291
; TYPE: PRT
; ORGANISM: Protein Multigene Family
US-09-846-808-12

Query Match      5.8%; Score 17; DB 9; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      249 VMGNQFKNLNVQHVHVEL 265
Db      256 VMGNQFKNLNVQHVHVEL 272
|||||

RESULT 19
US-10-059-964-50
; Sequence 50, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20020120115A1
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; CURRENT FILING DATE: 2002-01-28
; EARLIER APPLICATION NUMBER: 09/314,701
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn.Ver. 2.0
; SEQ ID NO 50
; LENGTH: 300
; TYPE: PRT
; ORGANISM: OMP-12
US-10-059-964-50

Query Match      5.8%; Score 17; DB 12; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      249 VMGNQFKNLNVQHVHVEL 265
Db      256 VMGNQFKNLNVQHVHVEL 272
|||||

RESULT 20
US-10-284-986-10
; Sequence 10, Application US/10284986
; Publication No. US20030091588A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xu-Jie
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane
; FILE REFERENCE: D6311D1
; CURRENT APPLICATION NUMBER: US/10/284,986
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: 09/846,808
; PRIOR FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 10
; LENGTH: 291
; TYPE: PRT
; ORGANISM: Protein Multigene Family
US-10-284-986-10
```



```
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: P28-10 Outer Membrane Protein of
; OTHER INFORMATION: Ehrlichia chaffeensis
US-10-284-986-10

Query Match          5.5%; Score 16; DB 9; Length 291;
Best Local Similarity 100.0%; Pred. No. 1.7e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 GLYISGQKPSVSVFS 53
Db 37 GLYISGQKPSVSVFS 52

RESULT 21
US-10-314-639-18
; Sequence 18, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; PRIOR FILING DATE: 2002-12-09
; PRIOR FILING DATE: 2002-12-09
; PRIOR FILING DATE: 2002-12-09
; PRIOR FILING DATE: 2002-12-09
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; TYPE: PRT
; LENGTH: 291
; ORGANISM: Ehrlichia chaffeensis
US-10-314-639-18

Query Match          5.5%; Score 16; DB 9; Length 291;
Best Local Similarity 100.0%; Pred. No. 1.7e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 GLYISGQKPSVSVFS 53
Db 37 GLYISGQKPSVSVFS 52

RESULT 22
US-09-846-808-10
; Sequence 10, Application US/09846808
; Patent No. US20020064531A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xu-Jie
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane
; FILE REFERENCE: D6311
; CURRENT APPLICATION NUMBER: US/09/846,808
; CURRENT FILING DATE: 2001-05-01
; PRIOR FILING DATE: 2001-05-01
; PRIOR FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 10
; TYPE: PRT
; LENGTH: 291
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: P28-10 Outer Membrane Protein of
; OTHER INFORMATION: Ehrlichia chaffeensis
US-09-846-808-10

Query Match          5.5%; Score 16; DB 10; Length 291;
Best Local Similarity 100.0%; Pred. No. 1.7e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 GLYISGQKPSVSVFS 53
Db 37 GLYISGQKPSVSVFS 52

RESULT 23
US-10-059-964-18
; Sequence 18, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20020120115A1
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; CURRENT FILING DATE: 2002-01-28
; EARLIER APPLICATION NUMBER: 09/314,701
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; TYPE: PRT
; LENGTH: 291
; ORGANISM: Ehrlichia chaffeensis
US-10-059-964-18

Query Match          5.5%; Score 16; DB 12; Length 291;
Best Local Similarity 100.0%; Pred. No. 1.7e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 GLYISGQKPSVSVFS 53
Db 37 GLYISGQKPSVSVFS 52

RESULT 24
US-10-314-639-62
; Sequence 62, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20030103991A1
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; CURRENT FILING DATE: 2002-12-09
; PRIOR FILING DATE: 2002-12-09
; PRIOR FILING DATE: 2002-12-09
; PRIOR FILING DATE: 2002-12-09
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; TYPE: PRT
; LENGTH: 241
; ORGANISM: p30-12
US-10-314-639-62

Query Match          5.1%; Score 15; DB 9; Length 241;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 EIEGSYEEDVDKPNP 137
Db 74 EIEGSYEEDVDKPNP 88

RESULT 25
US-10-059-964-62
; Sequence 62, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
```

; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20020120115A1io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; TITLE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/059,964  
; CURRENT FILING DATE: 2002-01-28  
; EARLIER APPLICATION NUMBER: 09/314,701  
; EARLIER FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 62  
; LENGTH: 241  
; TYPE: PRT  
; ORGANISM: P30-12  
US-10-059-964-62

Query Match 5.1%; Score 15; DB 12; Length 241;  
Best Local Similarity 100.0%; Pred. No. 1.3e-06; Indels 0; Gaps 0;  
Matches 15; Conservative 0; Mismatches 0;

Qy 123 EIEGSYEEDVKNKP 137  
Db 74 EIEGSYEEDVKNKP 88

## RESULT 26

US-10-284-986-8  
; Sequence 8, Application US/10284986  
; Publication No. US20030091588A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: Yu, Xu-Jie  
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane  
; TITLE OF INVENTION: Protein Multigene Family  
; FILE REFERENCE: D6311D1  
; CURRENT APPLICATION NUMBER: US/10/284,986  
; CURRENT FILING DATE: 2002-10-31  
; PRIOR APPLICATION NUMBER: 09/846,808  
; PRIOR FILING DATE: 2001-05-01  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 8  
; LENGTH: 275  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: P28-8 Outer Membrane Protein of  
; OTHER INFORMATION: Ehrlichia chaffeensis  
US-10-284-986-8

Query Match 5.1%; Score 15; DB 9; Length 275;  
Best Local Similarity 100.0%; Pred. No. 1.5e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 274 AVATLNIGYFGGEIG 288  
Db 256 AVATLNIGYFGGEIG 270

## RESULT 27

US-10-314-639-28  
; Sequence 28, Application US/10314639  
; Publication No. US20030103991A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20030103991A1io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; TITLE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/314,639  
; CURRENT FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US/09/314,701  
; PRIOR FILING DATE: 1999-05-19

; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 28  
; LENGTH: 275  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
US-10-314-639-28

Query Match 5.1%; Score 15; DB 9; Length 275;  
Best Local Similarity 100.0%; Pred. No. 1.5e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 274 AVATLNIGYFGGEIG 288  
Db 256 AVATLNIGYFGGEIG 270

## RESULT 28

US-09-846-808-8  
; Sequence 8, Application US/09846808  
; Patent No. US20020064531A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: Yu, Xu-Jie  
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane  
; TITLE OF INVENTION: Protein Multigene Family  
; FILE REFERENCE: D6311  
; CURRENT APPLICATION NUMBER: US/09/846,808  
; CURRENT FILING DATE: 2001-05-01  
; PRIOR APPLICATION NUMBER: 60/201,035  
; PRIOR FILING DATE: 2000-05-01  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 8  
; LENGTH: 275  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: P28-8 Outer Membrane Protein of  
; OTHER INFORMATION: Ehrlichia chaffeensis  
US-09-846-808-8

Query Match 5.1%; Score 15; DB 10; Length 275;  
Best Local Similarity 100.0%; Pred. No. 1.5e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 274 AVATLNIGYFGGEIG 288  
Db 256 AVATLNIGYFGGEIG 270

## RESULT 29

US-10-059-964-28  
; Sequence 28, Application US/10059964  
; Patent No. US20020120115A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20020120115A1io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; TITLE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/059,964  
; CURRENT FILING DATE: 2002-01-28  
; EARLIER APPLICATION NUMBER: 09/314,701  
; EARLIER FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 28  
; LENGTH: 275  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
US-10-059-964-28

Query Match 5.1%; Score 15; DB 12; Length 275;

```
Best Local Similarity 100.0%; Pred. No. 1.5e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 274 AVATLNICYFGGEIG 288
Db 256 AVATLNICYFGGEIG 270

RESULT 30
US-10-314-639-58
; Sequence 58, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20030103991A1
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; CURRENT FILING DATE: 2002-12-09
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 58
; LENGTH: 281
; TYPE: PRT
; ORGANISM: p30-9
US-10-314-639-58

Query Match 4.4%; Score 13; DB 9; Length 281;
Best Local Similarity 100.0%; Pred. No. 0.00014;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 272 TSAVATLNICYFG 284
Db 260 TSAVATLNICYFG 272

RESULT 31
US-10-059-964-58
; Sequence 58, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20020120115A1
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; CURRENT FILING DATE: 2002-01-28
; EARLIER APPLICATION NUMBER: 09/314,701
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 58
; LENGTH: 281
; TYPE: PRT
; ORGANISM: p30-9
US-10-059-964-58

Query Match 4.4%; Score 13; DB 12; Length 281;
Best Local Similarity 100.0%; Pred. No. 0.00014;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 272 TSAVATLNICYFG 284
Db 260 TSAVATLNICYFG 272

RESULT 32
US-10-284-986-6
; Sequence 6, Application US/10284986

; Publication No. US20030091588A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xu-Jie
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane
; TITLE OF INVENTION: Protein Multigene Family
; FILE REFERENCE: D6311D1
; CURRENT APPLICATION NUMBER: US/10/284,986
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: 09/846,808
; PRIOR FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 6
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: P28-6 Outer Membrane Protein of
; OTHER INFORMATION: Ehrlichia chaffeensis
US-10-284-986-6

Query Match 3.8%; Score 11; DB 9; Length 279;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 52 FSNFSVKETNV 62
Db 53 FSNFSVKETNV 63

RESULT 33
US-10-314-639-24
; Sequence 24, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20030103991A1
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/314,701
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-10-314-639-24

Query Match 3.8%; Score 11; DB 9; Length 279;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 52 FSNFSVKETNV 62
Db 53 FSNFSVKETNV 63

RESULT 34
US-09-846-808-6
; Sequence 6, Application US/09846808
; Patent No. US20020064531A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xu-Jie
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane
; TITLE OF INVENTION: Protein Multigene Family
; FILE REFERENCE: D6311
; CURRENT APPLICATION NUMBER: US/09/846,808
; CURRENT FILING DATE: 2001-05-01
```

```
; PRIOR APPLICATION NUMBER: 60/201,035
; PRIOR FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 6
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: P28-6 Outer Membrane Protein of
; OTHER INFORMATION: Ehrlichia chaffeensis
US-09-846-808-6

Query Match          3.8%; Score 11; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 PSNFSVKETNV 62
DB 53 PSNFSVKETNV 63

RESULT 35
US-10-059-964-24
; Sequence 24, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20020120115A1io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; EARLIER FILING DATE: 2002-01-28
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-10-059-964-24

Query Match          3.8%; Score 11; DB 12; Length 279;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 PSNFSVKETNV 62
DB 53 PSNFSVKETNV 63

RESULT 36
US-10-284-986-7
; Sequence 7, Application US/10284986
; Publication No. US20030091568A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xu-Jie
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane
; FILE REFERENCE: D6311D1
; CURRENT APPLICATION NUMBER: US/10/284,986
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: 09/846,808
; PRIOR FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 7
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: P28-7 Outer Membrane Protein of
; OTHER INFORMATION: Ehrlichia chaffeensis
US-09-846-808-7

Query Match          3.8%; Score 11; DB 10; Length 283;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 SYEEFDVKNPG 137
DB 122 SYEEFDVKNPG 132

RESULT 37
US-10-314-639-26
; Sequence 26, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20030103991A1io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/314,701
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-10-314-639-26

Query Match          3.8%; Score 11; DB 9; Length 283;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 SYEEFDVKNPG 137
DB 122 SYEEFDVKNPG 132

RESULT 38
US-09-846-808-7
; Sequence 7, Application US/09846808
; Patent No. US20020064531A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xu-Jie
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane
; FILE REFERENCE: D6311
; CURRENT APPLICATION NUMBER: US/09/846,808
; CURRENT FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 60/201,035
; PRIOR FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 7
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: P28-7 Outer Membrane Protein of
; OTHER INFORMATION: Ehrlichia chaffeensis
US-09-846-808-7

Query Match          3.8%; Score 11; DB 10; Length 283;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 SYEEFDVKNPG 137
DB 122 SYEEFDVKNPG 132
```

Db 122 SYEEFDVKVNG 132

## RESULT 39

US-10-059-964-26  
; Sequence 26, Application US/10059964  
; Patent No. US20020120115A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20020120115A1io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; TITLE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/059,964  
; CURRENT FILING DATE: 2002-01-28  
; EARLIER APPLICATION NUMBER: 09/314,701  
; EARLIER FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 26  
; LENGTH: 283  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
US-10-059-964-26

Query Match 3.8%; Score 11; DB 12; Length 283;  
Best Local Similarity 100.0%; Pred.No. 0.013;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 SYEEFDVKVNG 137

Db 122 SYEEFDVKVNG 132

## RESULT 40

US-10-284-986-5  
; Sequence 5, Application US/10284986  
; Publication No. US20030091588A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: Yu, Xu-Jie  
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane  
; TITLE OF INVENTION: Protein Multigene Family  
; FILE REFERENCE: D6311D1  
; CURRENT APPLICATION NUMBER: US/10/284,986  
; CURRENT FILING DATE: 2002-10-31  
; PRIOR APPLICATION NUMBER: 09/846,808  
; PRIOR FILING DATE: 2001-05-01  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 5  
; LENGTH: 295  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: P28-5 Outer Membrane Protein of  
; OTHER INFORMATION: Ehrlichia chaffeensis  
US-10-284-986-5

Query Match 3.8%; Score 11; DB 9; Length 295;  
Best Local Similarity 100.0%; Pred.No. 0.014;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 GLYISGQYKPS 48

Db 38 GLYISGQYKPS 48

Search completed: July 8, 2003, 10:42:34  
Job time : 73 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 10:28:26 ; Search time 40 Seconds  
(without alignments)  
704.185 Million cell updates/sec

Title: US-10-062-624-40  
Perfect score: 293  
Sequence: 1 MNVKLFKIITVLCVLLSL.....AVATLNIGYFGGIGARLTF 293

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0.

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR.73.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	2.7	166	2	G81235
2	8	2.7	353	2	T36274
3	8	2.7	620	2	AP2174
4	8	2.7	1036	2	G83329
5	8	2.7	1324	2	T01508
6	7	2.4	93	2	T23016
7	7	2.4	146	2	B95418
8	7	2.4	196	2	S58350
9	7	2.4	197	2	D69396
10	7	2.4	244	2	B83142
11	7	2.4	253	2	T04642
12	7	2.4	259	2	G70635
13	7	2.4	292	2	D64782
14	7	2.4	292	2	B90700
15	7	2.4	292	2	E85550
16	7	2.4	292	2	AB0557
17	7	2.4	293	2	AD3495
18	7	2.4	303	2	H81395
19	7	2.4	310	2	S58090
20	7	2.4	312	2	AH0634
21	7	2.4	312	2	AH2476
22	7	2.4	322	2	D97349
23	7	2.4	334	2	T14615
24	7	2.4	334	2	F86787
25	7	2.4	343	2	T31269
26	7	2.4	351	2	T25622
27	7	2.4	358	2	C69348
28	7	2.4	376	2	T24223
29	7	2.4	384	2	B82973

30	7	2.4	388	2	C83102	probable MFS trans
31	7	2.4	433	2	AF1739	PTS system, cellob
32	7	2.4	455	2	H97237	membrane associate
33	7	2.4	479	2	A61368	keratin type II, h
34	7	2.4	491	2	S05408	keratin, type II,
35	7	2.4	493	2	E69487	protein translocas
36	7	2.4	495	2	I57463	keratin type II -
37	7	2.4	506	2	T34712	probable hydrolase
38	7	2.4	508	1	KRSHL2	keratin type II, m
39	7	2.4	513	2	F72619	probable NADH dehy
40	7	2.4	594	2	S33561	ref(2)P protein -
41	7	2.4	604	2	S54032	probable amino aci
42	7	2.4	627	2	AB0535	hypothetical prote
43	7	2.4	628	2	S37795	hypothetical prote
44	7	2.4	629	2	A29666	keratin, 65K type
45	7	2.4	762	2	T50155	hypothetical prote
46	7	2.4	783	2	E97340	uncharacterized co
47	7	2.4	827	2	C84546	probable transport
48	7	2.4	829	2	T19494	hypothetical prote
49	7	2.4	851	2	S74442	ferric aerobactin
50	7	2.4	863	2	B64138	uridylyltransferas
51	7	2.4	890	2	S44150	coat protein - str
52	7	2.4	900	1	S25322	bifunctional beta-
53	7	2.4	1024	2	AB0347	AcrtB/AcrB/AcrF fam
54	7	2.4	1025	2	C64974	hypothetical prote
55	7	2.4	1025	2	D90989	hypothetical prote
56	7	2.4	1025	2	G85834	hypothetical prote
57	7	2.4	1026	2	AE0771	probable RND-famil
58	7	2.4	1046	2	A86790	ATP-dependent dsDN
59	7	2.4	1050	2	G70396	cation efflux syst
60	7	2.4	1099	2	T16283	hypothetical prote
61	7	2.4	1109	2	B90127	DNA-directed RNA p
62	7	2.4	1213	2	T37959	hypothetical prote
63	7	2.4	1487	2	S48719	phospholipase-A(2)
64	7	2.4	1696	2	T00057	hypothetical prote
65	7	2.4	3587	2	I40486	surfactin syntheta
66	6	2.0	38	2	D90631	hypothetical prote
67	6	2.0	46	2	S20018	hypothetical prote
68	6	2.0	50	2	PD0022	protein kinase (EC
69	6	2.0	74	2	G83393	hypothetical prote
70	6	2.0	79	2	G82709	hypothetical prote
71	6	2.0	80	2	S69784	p9gi protein - Sal
72	6	2.0	80	2	G89792	conserved hypothet
73	6	2.0	80	2	AE0849	pathogenicity 1 is
74	6	2.0	87	2	T17936	hypothetical prote
75	6	2.0	88	2	T07568	ribosomal protein
76	6	2.0	88	2	AC0730	hypothetical prote
77	6	2.0	94	2	S77847	transposase ssr117
78	6	2.0	96	2	F69516	hypothetical prote
79	6	2.0	97	2	T29766	hypothetical prote
80	6	2.0	99	2	I50100	gastin-releasing
81	6	2.0	100	2	T09148	late-embryogenesis
82	6	2.0	100	2	S69457	hypothetical prote
83	6	2.0	102	2	D84090	hypothetical prote
84	6	2.0	107	2	B69262	hypothetical prote
85	6	2.0	108	1	K1H0HU	Ig kappa chain V-I
86	6	2.0	108	2	S75384	hypothetical prote
87	6	2.0	109	1	K3HUPM	Ig kappa chain V-I
88	6	2.0	111	2	C72805	gp44.i protein - M
89	6	2.0	111	2	T17730	hypothetical prote
90	6	2.0	113	2	B54437	apoptotic cell dea
91	6	2.0	114	2	S00996	Ig kappa chain pre
92	6	2.0	114	2	A82983	conserved hypothet
93	6	2.0	115	2	B75535	ribosomal protein
94	6	2.0	116	2	I34284	NADH2 dehydrogenas
95	6	2.0	116	2	AC1571	Salmonelle enteric
96	6	2.0	116	2	AH1217	Salmonelle enteric
97	6	2.0	117	2	B82646	conserved hypothet
98	6	2.0	118	2	AE2234	50S ribosomal prot
99	6	2.0	118	2	C72490	hypothetical prote
100	6	2.0	119	2	S63356	probable membrane
101	6	2.0	119	2	S74925	transposase slr1065
102	6	2.0	119	2	S75488	transposase slr211

103	6	2.0	119	2	S74836	transposase slr085	176	6	2.0	183	2	T04217	hypothetical prote
104	6	2.0	119	2	S75590	transposase slr125	177	6	2.0	184	2	S65918	RNA-directed DNA p
105	6	2.0	119	2	S75590	minor pilin - Aqu	178	6	2.0	185	2	A97118	probable sigma fac
106	6	2.0	120	2	F70424	hypothetical prote	179	6	2.0	186	2	A99975	ica operon transcrip
107	6	2.0	123	2	AH2707	conserved hypoteth	180	6	2.0	187	2	E70873	hypothetical prote
108	6	2.0	125	2	S40353	Ig kappa chain V-J	181	6	2.0	187	2	C71317	hypothetical prote
109	6	2.0	126	2	S39903	hypothetical prote	182	6	2.0	189	2	T19778	hypothetical prote
110	6	2.0	127	2	B81387	hypothetical prote	183	6	2.0	189	2	S48914	conserved hypoteth
111	6	2.0	128	2	C81056	conserved hypoteth	184	6	2.0	190	2	H69192	probable heat shoc
112	6	2.0	128	2	E72804	gp39 protein - Myc	185	6	2.0	192	2	D81184	hypothetical prote
113	6	2.0	130	2	F64692	hypothetical prote	186	6	2.0	196	2	E84807	hypothetical prote
114	6	2.0	132	2	B83379	hypothetical prote	187	6	2.0	196	2	S19408	hypothetical prote
115	6	2.0	133	2	C65135	ribosome-associate	188	6	2.0	197	2	S51464	probable membrane
116	6	2.0	133	2	C65135	ribosome-associate	189	6	2.0	198	2	T41529	hypothetical prote
117	6	2.0	133	2	B91159	heat shock protein	190	6	2.0	199	2	T26949	hypothetical prote
118	6	2.0	133	2	A10998	ribosome-associate	191	6	2.0	199	2	E71718	hypothetical prote
119	6	2.0	133	2	H86004	hypothetical prote	192	6	2.0	200	2	A26169	nucleoplasmin - Af
120	6	2.0	133	2	B48234	hypothetical prote	193	6	2.0	200	2	B64621	conserved hypoteth
121	6	2.0	135	2	JH0684	interleukin-4 prec	194	6	2.0	200	2	A71895	hypothetical prote
122	6	2.0	135	2	I46142	interleukin 4 - re	195	6	2.0	200	2	G97160	general stress pro
123	6	2.0	136	2	A91337	Ig kappa chain pre	196	6	2.0	200	2	E95312	hypothetical prote
124	6	2.0	136	2	S12098	rfau protein - Sal	197	6	2.0	201	2	DAPSAA	protocatechuate 3,
125	6	2.0	137	2	A11375	methylmalonyl-CoA	198	6	2.0	201	2	F91295	hyperosmotically i
126	6	2.0	140	2	C84407	hypothetical prote	199	6	2.0	201	2	AC0363	conserved hypoteth
127	6	2.0	143	2	G97166	flagellar basal bo	200	6	2.0	201	2	A41899	hyperosmotically i
128	6	2.0	144	2	D64115	transcription term	201	6	2.0	203	2	S09559	hypothetical prote
129	6	2.0	144	2	AC2572	hypothetical prote	202	6	2.0	203	2	T32745	hypothetical prote
130	6	2.0	145	1	A64214	hypothetical prote	203	6	2.0	205	2	AE1072	Putative periplasm
131	6	2.0	145	2	F01068	flagellar prote	204	6	2.0	205	2	S76688	2-dehydro-3-deoxy-
132	6	2.0	145	2	AE2637	conserved hypoteth	205	6	2.0	206	2	C85482	hypothetical prote
133	6	2.0	146	1	A71421	hypothetical 15.5K	206	6	2.0	207	2	C86147	TiN6.2 protein - A
134	6	2.0	146	2	S15449	hemoglobin beta ch	207	6	2.0	208	2	E97094	ncharacterized con
135	6	2.0	147	2	D69339	conserved hypoteth	208	6	2.0	209	2	F70159	ribosomal protein
136	6	2.0	147	2	C95274	hypothetical prote	209	6	2.0	209	2	C89005	protein f24A6.3 [i
137	6	2.0	149	2	T11066	NADH2 dehydrogenas	210	6	2.0	210	2	C87256	hypothetical prote
138	6	2.0	149	2	C90824	hypothetical membr	211	6	2.0	211	2	D65011	hypothetical prote
139	6	2.0	151	2	T19087	hypothetical prote	212	6	2.0	211	2	G91035	hypothetical prote
140	6	2.0	153	2	F97269	dihydrofolate redu	213	6	2.0	211	2	A85880	amidase related to
141	6	2.0	153	2	A60585	thyroglobulin - sl	214	6	2.0	211	2	H96995	hypothetical prote
142	6	2.0	156	2	H95252	PTS system, IIB co	215	6	2.0	212	2	T39489	hypothetical prote
143	6	2.0	156	2	F98117	hypothetical prote	216	6	2.0	213	1	S58329	HGH49 protein - ye
144	6	2.0	157	2	AG3560	transcription regu	217	6	2.0	213	2	S74247	CPD diacylglycerol
145	6	2.0	160	2	S73841	dihydrofolate redu	218	6	2.0	213	2	T50724	hypothetical prote
146	6	2.0	160	2	T21504	hypothetical prote	219	6	2.0	215	1	LANT56	ribonuclease (EC 3
147	6	2.0	160	2	D97419	VGN1536C protein (	220	6	2.0	216	2	AG2591	peptide methionine
148	6	2.0	163	1	S21632	hemoglobin Z precu	221	6	2.0	216	2	T17693	hypothetical prote
149	6	2.0	163	2	E70187	lipopolysaccharide	222	6	2.0	216	2	C84786	hypothetical prote
150	6	2.0	164	2	S50188	phosphotransferase	223	6	2.0	216	2	AF3415	histidinol-phospha
151	6	2.0	165	2	B71432	hypothetical prote	224	6	2.0	216	2	T27331	hypothetical prote
152	6	2.0	166	2	F89919	hypothetical prote	225	6	2.0	217	2	H59060	chromosome partiti
153	6	2.0	167	2	T15950	hypothetical prote	226	6	2.0	218	2	H27177	hypothetical prote
154	6	2.0	168	2	F70148	ribosomal protein	227	6	2.0	220	2	F82574	hypothetical prote
155	6	2.0	168	2	B96916	hypothetical prote	228	6	2.0	220	2	D84420	hypothetical prote
156	6	2.0	168	2	G83674	phenylacetic acid	229	6	2.0	221	2	AD3321	trans-aconitate me
157	6	2.0	171	2	S75475	transposase slr152	230	6	2.0	221	2	H87496	hypothetical prote
158	6	2.0	171	2	S25724	hypothetical prote	231	6	2.0	222	2	H83387	probable hydrolase
159	6	2.0	173	2	B85875	probable minor fim	232	6	2.0	222	2	AE0182	ABC transporter Ar
160	6	2.0	173	2	A72321	conserved hypoteth	233	6	2.0	223	2	E84175	hypothetical prote
161	6	2.0	175	2	D71195	methylated-DNA-pro	234	6	2.0	223	2	T03238	extensin (clone Ex
162	6	2.0	175	2	C70326	hypothetical prote	235	6	2.0	224	2	A65113	hypothetical prote
163	6	2.0	175	2	AH1589	hypothetical prote	236	6	2.0	224	2	S41031	hypothetical prote
164	6	2.0	176	2	F96648	hypothetical prote	237	6	2.0	225	2	C96794	hypothetical prote
165	6	2.0	177	2	T28263	ORF MSV102 hypot	238	6	2.0	226	2	C82152	ABC transporter (A
166	6	2.0	178	2	E83025	conserved hypoteth	239	6	2.0	227	2	C82152	ABC transporter (A
167	6	2.0	179	2	S75214	hypothetical prote	240	6	2.0	228	2	A84088	ToIQ colicin impor
168	6	2.0	179	2	D90361	hypothetical prote	241	6	2.0	228	2	A10137	hypothetical prote
169	6	2.0	179	2	T22734	hypothetical prote	242	6	2.0	228	2	H97373	hypothetical prote
170	6	2.0	180	2	AF1228	B. subtilis yshB p	243	6	2.0	228	2	C75463	hypothetical prote
171	6	2.0	180	2	AH1581	B. subtilis yshB p	244	6	2.0	228	2	T49057	hypothetical prote
172	6	2.0	181	2	A40607	monofunctional cho	245	6	2.0	229	2	S73777	hypothetical prote
173	6	2.0	181	2	G84680	hypothetical prote	246	6	2.0	230	1	BVECTQ	biopolymer transpo
174	6	2.0	182	2	A91031	probable fibrillar-	247	6	2.0	230	2	S70601	H+-transporting tw
175	6	2.0	183	2	T26637	hypothetical prote	248	6	2.0	230	2	D90725	inner membrane pro



249	6	2.0	230	2	E85576	inner membrane pro	322	2.0	277	2	D64531	type II restrictio
250	6	2.0	230	2	A80592	TolQ protein [impo	323	2.0	277	2	H86589	hypothetical prote
251	6	2.0	230	2	A86445	unknown protein [i	324	2.0	277	2	C72033	hypothetical prote
252	6	2.0	230	2	G97144	hypothetical prote	325	2.0	278	2	S23519	beta-luffin - smoo
253	6	2.0	230	2	B95312	hypothetical prote	326	2.0	278	2	C70138	hypothetical prote
254	6	2.0	231	2	S57271	14-3-3 protein hom	327	2.0	278	2	AG2176	hypothetical prote
255	6	2.0	231	2	A81881	hypothetical prote	328	2.0	279	2	TS0125	probable 1-acylgly
256	6	2.0	233	2	S71463	ribonuclease (EC 3	329	2.0	279	2	AB1745	conserved hypotnet
257	6	2.0	235	2	D97715	hypothetical prote	330	2.0	280	2	AD1375	conserved hypotnet
258	6	2.0	235	2	AC3649	transcription acti	331	2.0	280	2	T03236	extensin precursor
259	6	2.0	236	2	A98351	agropine synthesis	332	2.0	280	2	H84532	hypothetical prote
260	6	2.0	236	2	A82931	agropine synthesis	333	2.0	281	2	S37745	yyfL protein - Bac
261	6	2.0	236	2	D59433	branched-chain ami	334	2.0	282	2	S76906	transposase sl1043
262	6	2.0	238	2	E90393	conserved hypotnet	335	2.0	282	2	S76382	transposase slr023
263	6	2.0	240	2	T24363	hypothetical prote	336	2.0	282	2	S76312	transposase slr035
264	6	2.0	240	2	C64152	hypothetical prote	337	2.0	282	2	S77237	transposase slr135
265	6	2.0	241	2	C90222	proteasome subunit	338	2.0	282	2	S75484	transposase sl1199
266	6	2.0	241	2	T26909	hypothetical prote	339	2.0	283	2	T30365	occlusion-derived
267	6	2.0	241	2	C45594	ORF 3' of calmodul	340	2.0	283	2	TI5964	hypothetical prote
268	6	2.0	244	1	A40668	hemolymph juvenile	341	2.0	285	2	JC6007	transcription acti
269	6	2.0	244	2	G90094	hypothetical prote	342	2.0	285	2	G86323	protein F14D16.15
270	6	2.0	245	2	AH1298	E. coli tRNA (guan	343	2.0	285	2	C83254	probable hydrolase
271	6	2.0	245	2	AH1670	E. coli tRNA (guan	344	2.0	287	2	B69148	hypothetical prote
272	6	2.0	247	2	S26733	hypothetical prote	345	2.0	288	2	S69661	hypothetical prote
273	6	2.0	248	2	G59984	rRNA methylase hom	346	2.0	289	2	AH2088	heat shock protein
274	6	2.0	249	2	T19088	hypothetical prote	347	2.0	289	2	B89592	protein TolB10.2 [
275	6	2.0	249	2	D87582	hypothetical prote	348	2.0	291	2	B71097	hypothetical prote
276	6	2.0	249	2	E86231	hypothetical prote	349	2.0	291	2	G71300	probable protein-m
277	6	2.0	249	2	S30580	U2 snRNP protein A	350	2.0	292	2	S25818	hypothetical prote
278	6	2.0	250	2	JN0108	luffin-b - smooth	351	2.0	293	2	B69727	translation elonga
279	6	2.0	250	2	T15887	hypothetical prote	352	2.0	293	2	B83953	elongation factor
280	6	2.0	250	2	T46417	hypothetical prote	353	2.0	293	2	D81673	ABC transporter, p
281	6	2.0	251	2	B69485	daunorubicin resis	354	2.0	293	2	S29384	cyclohexadienyl de
282	6	2.0	251	2	S62202	hypothetical prote	355	2.0	294	2	H64632	phosphate acetyltr
283	6	2.0	251	2	A55523	hypothetical prote	356	2.0	294	2	EA4084	two-component sens
284	6	2.0	252	2	G97886	hypothetical prote	357	2.0	295	2	AD1461	dimethyladenosine
285	6	2.0	252	2	A80876	conserved hypotnet	358	2.0	295	2	AE1098	dimethyladenosine
286	6	2.0	254	2	S16801	1-(5-phosphoribosy	359	2.0	295	2	G69202	hypothetical prote
287	6	2.0	255	2	AB3280	membrane-bound lyt	360	2.0	296	2	D89936	hypothetical prote
288	6	2.0	256	1	C69304	probable enoyl-CoA	361	2.0	299	2	T06960	probable membrane
289	6	2.0	257	1	S74028	sulfate transport	362	2.0	300	1	H70321	conserved hypotnet
290	6	2.0	257	2	E82871	dimethyladenosine	363	2.0	301	2	E96921	oligopeptide trans
291	6	2.0	257	2	AC2406	2-hydroxyhepta-2,4	364	2.0	301	2	A83017	probable transcrip
292	6	2.0	260	2	E97310	pyrroline-5-carbox	365	2.0	302	2	D69118	heterodisulfide re
293	6	2.0	260	2	T07388	14-3-3 protein tft	366	2.0	302	2	S78509	heterodisulfide re
294	6	2.0	260	2	T02051	14-3-3 protein hom	367	2.0	303	2	H89881	hypothetical prote
295	6	2.0	260	2	E97203	hypothetical prote	368	2.0	303	2	F84220	hypothetical prote
296	6	2.0	261	2	D71523	probable outer mem	369	2.0	303	2	T03571	citrate (pro-38)-l
297	6	2.0	262	2	T48546	hypothetical prote	370	2.0	303	2	B47089	probable peptidase
298	6	2.0	263	2	E81340	flagellar basal-bo	371	2.0	304	2	D96952	probable AraC-type
299	6	2.0	264	2	C64701	hypothetical prote	372	2.0	304	2	E90521	glycerol-3-phospha
300	6	2.0	265	2	C83419	NADH-dependent eno	373	2.0	304	2	H90525	hypothetical prote
301	6	2.0	265	2	H96777	hypothetical prote	374	2.0	305	2	D90266	conserved hypotnet
302	6	2.0	266	2	F64977	Phosphomethylpyrim	375	2.0	307	2	S50586	hypothetical prote
303	6	2.0	266	2	B90992	phosphomethylpyrim	376	2.0	307	2	AC2364	hypothetical prote
304	6	2.0	266	2	D85837	phosphomethylpyrim	377	2.0	308	2	JC2253	chitinase (EC 3.2.
305	6	2.0	267	1	A64745	probable aldehyde	378	2.0	309	2	S51573	mobB protein precu
306	6	2.0	267	2	C90654	2,5-diketo-D-gluc	379	2.0	309	2	C96019	probable rhizopine
307	6	2.0	267	2	C85505	aldehyde reductase	380	2.0	309	2	G90600	hpr(ser) kinase/ph
308	6	2.0	267	2	A80533	hypothetical oxido	381	2.0	310	1	KIECM	homoserine kinase
309	6	2.0	268	2	D84223	probable 1-acylgly	382	2.0	310	2	C90629	homoserine kinase
310	6	2.0	268	2	T22226	hypothetical prote	383	2.0	310	2	C85480	homoserine kinase
311	6	2.0	269	2	J70525	tryptophan synthas	384	2.0	310	2	D84905	hypothetical prote
312	6	2.0	269	2	T15635	hypothetical prote	385	2.0	310	2	S43928	estradiol 17beta-d
313	6	2.0	269	2	S77394	hypothetical prote	386	2.0	310	2	D72615	hypothetical prote
314	6	2.0	270	2	AH2435	hypothetical prote	387	2.0	312	2	S48849	chalcone reductase
315	6	2.0	272	2	JC4811	betavulgin - beet	388	2.0	312	2	S48850	chalcone reductase
316	6	2.0	272	2	E97658	hypothetical prote	389	2.0	312	2	S48851	chalcone reductase
317	6	2.0	272	2	AF2882	conserved hypotnet	390	2.0	312	2	E83740	hypothetical prote
318	6	2.0	273	2	PC4153	monophenol monooxy	391	2.0	313	2	AC3116	ABC transporter, a
319	6	2.0	273	2	H82286	phosphate ABC tran	392	2.0	314	2	H84677	hypothetical prote
320	6	2.0	276	2	H75588	conserved hypotnet	393	2.0	315	1	SI4222	chalcone reductase
321	6	2.0	277	2	G71976	type II DNA modifi	394	2.0	315	2	AE3439	pirin [imported] -

395 2.0 316 2 AD1636  
396 2.0 317 2 H61245  
397 2.0 318 2 AC1858  
398 2.0 319 2 S14948  
399 2.0 320 2 S13158  
400 2.0 321 2 T06641  
401 2.0 322 1 E84026  
402 2.0 323 2 T27521  
403 2.0 324 1 PAECS  
404 2.0 325 2 H70617  
405 2.0 326 2 F82022  
406 2.0 327 2 D86138  
407 2.0 328 2 B91297  
408 2.0 329 2 D82435  
409 2.0 330 2 B84452  
410 2.0 331 2 S60335  
411 2.0 332 2 T05355  
412 2.0 333 2 G75117  
413 2.0 334 2 AE1384  
414 2.0 335 2 S40268  
415 2.0 336 2 G88115  
416 2.0 337 2 T31704  
417 2.0 338 2 B98171  
418 2.0 339 2 G84981  
419 2.0 340 2 AG3438  
420 2.0 341 2 H82340  
421 2.0 342 2 T21182  
422 2.0 343 2 AC2386  
423 2.0 344 2 H64566  
424 2.0 345 2 G90492  
425 2.0 346 2 A71860  
426 2.0 347 2 T33428  
427 2.0 348 2 A10972  
428 2.0 349 2 T16850  
429 2.0 350 2 C86921  
430 2.0 351 2 A88957  
431 2.0 352 2 D90982  
432 2.0 353 2 B85828  
433 2.0 354 2 S50238  
434 2.0 355 2 B64968  
435 2.0 356 2 D87801  
436 2.0 357 2 C72301  
437 2.0 358 2 E71918  
438 2.0 359 2 G86579  
439 2.0 360 2 B72043  
440 2.0 361 2 D64595  
441 2.0 362 2 AB3470  
442 2.0 363 2 C71266  
443 2.0 364 2 T46572  
444 2.0 365 2 AH0087  
445 2.0 366 2 T23500  
446 2.0 367 2 AG1851  
447 2.0 368 2 D70126  
448 2.0 369 2 T34981  
449 2.0 370 2 T31658  
450 2.0 371 2 T33706  
451 2.0 372 2 G83441  
452 2.0 373 2 T28213  
453 2.0 374 2 T33365  
454 2.0 375 2 T15079  
455 2.0 376 2 S10571  
456 2.0 377 2 T04184  
457 2.0 378 2 D84798  
458 2.0 379 1 S77070  
459 2.0 380 2 S72587  
460 2.0 381 2 G96584  
461 2.0 382 2 AE1694  
462 2.0 383 2 D86562  
463 2.0 384 1 MFNZB3  
464 2.0 385 2 S25480  
465 2.0 386 2 S71581  
466 2.0 387 2 C85909  
467 2.0 388 2

ornithine carbanoy  
glycerate dehydro  
melanocortin recep  
hypothetical prote  
chitinase (EC 3.2.  
extensin - common  
hypothetical prote  
probable S-adenosy  
hypothetical prote  
phosphoserine phos  
hypothetical prote  
probable glycerate  
phosphoserine phos  
3-phosphoserine ph  
D-3-phosphoglycer  
probable steroid s  
TGF-beta receptor  
hypothetical prote  
dipeptide abc tran  
UDP-glucose 4-epim  
peroxidase (EC 1.1  
protein F53C3.1 [i  
hypothetical prote  
probable rhizopine  
DNA-directed DNA p  
peptidoglycan bind  
gluconate utilizat  
hypothetical prote  
chaperone DnaJ pro  
ferrochelatase [EC  
hypothetical prote  
ferrochelatase - H  
hypothetical prote  
lipopolysaccharide  
hypothetical prote  
conserved hypoteth  
protein ZK697.10 [i  
hypothetical prote  
hypothetical prote  
vanadate resistanc  
O-polysaccharide a  
protein C10G11.7 [i  
endoglucanase - Th  
uroporphyrinogen d  
glutamyl tRNA redu  
glutamyl-tRNA redu  
uroporphyrinogen d  
glucose-resistance  
conserved hypoteth  
probable UDP-glucu  
sigma-54 transcrip  
hypothetical prote  
hypothetical prote  
hypothetical prote  
probable integral  
hypothetical prote  
conserved hypoteth  
DNA-directed RNA p  
hypothetical prote  
hypothetical prote  
mucin 1 precursor,  
hypothetical prote  
probable mitochond  
pyridoxal phosphat  
phoH protein - Myc  
hypothetical prote  
3-isopropylmalate  
peptide chain rele  
matrix protein - p  
heat shock transcr  
1-aminocyclopropan  
hypothetical prote

352 2 G91064  
352 2 B84524  
353 2 C70155  
354 2 T10474  
355 2 T37202  
356 2 A44869  
357 2 A11393  
358 2 E87039  
359 2 G70869  
360 2 A86615  
361 2 D72009  
362 2 T09263  
363 2 D81345  
364 2 T33756  
365 2 H70147  
366 2 JC2443  
367 2 F97631  
368 2 A12854  
369 2 E70892  
370 2 F83203  
371 2 AB0134  
372 2 B95371  
373 2 B89907  
374 2 T00200  
375 2 G86340  
376 2 E72475  
377 2 AB0792  
378 2 C84236  
379 2 T00580  
380 2 T23825  
381 2 D84846  
382 2 AB0276  
383 2 T08273  
384 2 T06276  
385 2 I38450  
386 2 F98308  
387 2 AE2974  
388 2 S63442  
389 2 D96037  
390 2 C82642  
391 2 G84404  
392 2 G85975  
393 2 B71564  
394 2 J70744  
395 2 JC5542  
396 2 B86712  
397 2 B86773  
398 2 S66728  
399 2 A75328  
400 2 JQ1737  
401 2 T11776  
402 2 I57682  
403 2 A41007  
404 2 I61769  
405 2 G91130  
406 2 T04914  
407 2 H86870  
408 2 T39588  
409 2 T23887  
410 2 A60093  
411 2 D70849  
412 2 AG0275  
413 2 B69252  
414 2 A87572  
415 2 JQ0410  
416 2 A64986  
417 2 AG0783  
418 2 F91011  
419 2 H85855  
420 2 E96954  
421 2 T03313  
422 2 D65167  
423 2 AE3356

hypothetical prote  
hypothetical prote  
nucleotide sugar e  
thiamin biosynthes  
hypothetical prote  
cell type-specific  
conserved hypoteth  
uroporphyrinogen d  
probable uroporph  
C7839 hypoteth  
conserved hypoteth  
malate dehydrogena  
probable periplasm  
hypothetical prote  
chemokine (C-C) re  
basic membrane pro  
hypothetical prote  
conserved hypoteth  
hypothetical prote  
probable ATP-bind  
probable regulator  
probable oxidoredu  
hypothetical prote  
hypothetical prote  
protein F2D10.35 [i  
hypothetical prote  
glutamate 5-kinase  
hypothetical prote  
probable lacyl-car  
hypothetical prote  
hypothetical prote  
probable membrane  
conserved hypoteth  
benzothiadiazole-1  
chemokine (C-C) re  
acetylornithine de  
probable ABC trans  
hypothetical prote  
hypothetical prote  
hypothetical prote  
probable coproporp  
site-specific DNA-  
DNA (cytosine-5')-  
hypothetical prote  
malate dehydrogena  
phoH-related prote  
hypothetical 42.2K  
ubiquinol-cytochro  
bombesin/ GRP rece  
gastrin-releasing  
keratin 6d, type I  
N-acetylgalactosam  
hypothetical prote  
hypothetical prote  
hypothetical prote  
cytokeratin, type  
probable aminotran  
mannose-6-phosphat  
3-ketoacyl-CoA thi  
aminotransferase,  
methionine adenosy  
probable membrane  
sugar efflux trans  
sugar efflux trans  
probable transport  
Na/H antiporter (n  
gene 16 protein -  
probable membrane  
lipid-A-disacchari

541	6	2.0	396	1	A42919	mevalonate kinase	614	2	C85618	mukF protein (kill
542	6	2.0	396	2	AE1796	efflux protein hom	615	2	AF0615	killing factor kic
543	6	2.0	397	2	G90048	hypothetical prote	616	2	H85911	hypothetical prote
544	6	2.0	398	1	C89971	probable phosphoe	617	6	T03172	helicase homolog 0
545	6	2.0	399	2	A46632	bombesin-like pept	618	6	T22632	hypothetical prote
546	6	2.0	399	2	S29480	bombesin receptor	619	6	D64161	hypothetical prote
547	6	2.0	399	2	E86197	hypothetical prote	620	6	G64234	hypothetical prote
548	6	2.0	399	2	T16634	hypothetical prote	621	6	G64234	hypothetical prote
549	6	2.0	400	2	A72689	probable nonspecif	622	6	S59771	hypothetical prote
550	6	2.0	402	2	C84529	probable polygalac	623	6	B96498	probable polygalac
551	6	2.0	402	2	D84529	probable polygalac	624	6	C96498	probable polygalac
552	6	2.0	402	2	T00995	probable polygalac	625	6	E95065	conserved hypothet
553	6	2.0	402	2	A11594	Na+/H+-exchangin	626	6	E97932	conserved hypothet
554	6	2.0	403	2	C70385	hypothetical prote	627	6	A12724	potassium uptake p
555	6	2.0	404	2	A36498	probable polygalac	628	6	F82166	mukF protein VCI71
556	6	2.0	404	2	H84827	probable polygalac	629	6	B40590	flagellar distal c
557	6	2.0	404	2	B84529	probable polygalac	630	6	AC3349	protein translocas
558	6	2.0	404	2	S65991	membrane protein y	631	6	F70302	transporter (Pho87
559	6	2.0	406	2	E82041	general secretion	632	6	A10442	probable exported
560	6	2.0	408	2	T08400	late embryonic abu	633	6	F71351	probable preprotei
561	6	2.0	408	2	T39570	probable metal tra	634	6	D83791	biotin carboxylase
562	6	2.0	409	2	A71493	hypothetical prote	635	6	B75006	DNA repair protein
563	6	2.0	409	2	E86441	46.3K hypothetical	636	6	A69763	homoserine dehydro
564	6	2.0	410	2	E84998	hypothetical prote	637	6	D91141	probable C4-dicarb
565	6	2.0	410	2	T04372	protein BEA1 - bar	638	6	G85986	probable transport
566	6	2.0	410	2	T06213	hypothetical prote	639	6	C72541	probable 3-phospho
567	6	2.0	411	2	T22290	hypothetical prote	640	6	T02593	hypothetical prote
568	6	2.0	412	2	D69525	arylsulfatase regu	641	6	C84133	glutamine syntheta
569	6	2.0	412	2	E65146	hypothetical 47.4	642	6	E65114	Putative cryptic C
570	6	2.0	413	2	S51667	serine/threonine k	643	6	S78296	conserved hypothet
571	6	2.0	413	2	AE0089	probable flagellar	644	6	S50725	hypothetical prote
572	6	2.0	417	2	F64915	membrane protein y	645	6	S70725	sodium,sulfate sym
573	6	2.0	417	2	F90916	probable transport	646	6	AF0093	probable phosphodi
574	6	2.0	417	2	C85765	probable transport	647	6	T09932	hypothetical prote
575	6	2.0	417	2	AB0682	probable membrane	648	6	B86600	D-alanine/glycine
576	6	2.0	418	2	A97300	gamma-glutamyl pho	649	6	B72024	hexose phosphate t
577	6	2.0	418	2	T52071	mRNA-binding prote	650	6	A89785	glycolate oxidase
578	6	2.0	419	1	KXKL2B	keratin, 64K type	651	6	H71528	hypothetical prote
579	6	2.0	419	2	T15088	hypothetical prote	652	6	A43782	keratin, type II -
580	6	2.0	419	2	S47843	rfal protein - Esc	653	6	A56239	mannose-6-phosphat
581	6	2.0	420	2	T41870	ALK-EXO orf133 - B	654	6	I64080	probable membrane
582	6	2.0	423	2	T33667	hypothetical prote	655	6	F90568	hypothetical prote
583	6	2.0	423	2	S74278	hypothetical prote	656	6	T33747	hypothetical prote
584	6	2.0	425	1	KXKL2A	hypothetical prote	657	6	P2WLC1	L2 protein - pygmy
585	6	2.0	426	2	AE1304	dihydroorotase hom	658	6	delta-2-crystallin	delta-2-crystallin
586	6	2.0	426	2	A82660	glucose/galactose	659	6	delta-1-crystallin	delta-1-crystallin
587	6	2.0	426	2	E84689	probable RING zinc	660	6	fumarate hydratase	fumarate hydratase
588	6	2.0	426	2	E71675	hypothetical prote	661	6	fumarase C import	fumarase C import
589	6	2.0	428	2	T31746	hypothetical prote	662	6	fumarase C import	fumarase C import
590	6	2.0	429	2	AE0346	4-aminobutyrate tr	663	6	fumarate hydratase	fumarate hydratase
591	6	2.0	429	2	C84728	hypothetical prote	664	6	phosphogluconate d	phosphogluconate d
592	6	2.0	431	2	S37775	filamin, muscle -	665	6	probable polygalac	probable polygalac
593	6	2.0	433	2	AD0161	serine transporter	666	6	glycine hydroxymet	glycine hydroxymet
594	6	2.0	433	2	E81304	signal-transductio	667	6	keratin K7, type I	keratin K7, type I
595	6	2.0	433	2	A75125	hypothetical prote	668	6	probable integrat	probable integrat
596	6	2.0	434	2	AB1601	E. coli YbdN prote	669	6	glutamyl-tRNA (gln)	glutamyl-tRNA (gln)
597	6	2.0	434	2	D65132	hypothetical 46.5	670	6	glutamyl-tRNA (gln)	glutamyl-tRNA (gln)
598	6	2.0	436	2	C95236	membrane protein l	671	6	hypothetical prote	hypothetical prote
599	6	2.0	436	2	D98100	conserved hypothet	672	6	hypothetical prote	hypothetical prote
600	6	2.0	437	2	A95202	rRNA methyltransfe	673	6	probable replicati	probable replicati
601	6	2.0	437	2	G91067	hypothetical prote	674	6	replicative DNA he	replicative DNA he
602	6	2.0	438	2	A64147	hypothetical prote	675	6	hypothetical prote	hypothetical prote
603	6	2.0	438	2	A98069	rRNA methylase lim	676	6	trbi protein - Ent	trbi protein - Ent
604	6	2.0	438	2	S76347	UDP-N-acetylglucos	677	6	succinate-semialde	succinate-semialde
605	6	2.0	438	2	B71963	probable outer mem	678	6	transposase - Bac	transposase - Bac
606	6	2.0	439	2	T43813	translation elonga	679	6	ABC transporter lim	ABC transporter lim
607	6	2.0	440	2	S47743	probable sugar tra	680	6	glutamyl-tRNA (gln	glutamyl-tRNA (gln
608	6	2.0	440	2	C91179	probable transport	681	6	sensor protein kin	sensor protein kin
609	6	2.0	440	2	D86025	probable transport	682	6		
610	6	2.0	440	2	AD0986	hypothetical metab	683	6		
611	6	2.0	440	2	T24478	hypothetical prote	684	6		
612	6	2.0	440	2	S43911	mukF protein - Esc	685	6		
613	6	2.0	440	2	E90754	mukF protein [impo	686	6		

687	6	2.0	482	2	S31478	alpha-amylase (EC	760	2.0	514	2	D97921	peptide chain rele
688	6	2.0	482	2	S01153	cell division cont	761	2.0	515	1	A32931	glucosylceramidase
689	6	2.0	482	2	A34948	cyclin-related cel	762	2.0	517	1	FWEMLA	legumin A precurs
690	6	2.0	482	2	D72559	hypothetical prote	763	2.0	517	2	D84421	probable amino aci
691	6	2.0	485	2	E97506	potassium uptake p	764	2.0	518	2	H97320	PTS system, (possi
692	6	2.0	486	2	A46341	helper component p	765	2.0	518	2	A64130	hypothetical prote
693	6	2.0	486	2	B46341	helper component p	766	2.0	520	2	A35883	laccase (EC 1.10.3
694	6	2.0	487	2	JT0407	keratin 8, type II	767	2.0	520	2	B35883	ligninolytic pheno
695	6	2.0	487	2	D64551	outer membrane pro	768	2.0	520	2	S08237	legumin A2 precurs
696	6	2.0	488	2	AE1419	IMP dehydrogenase	769	2.0	522	2	A46103	transmembrane glyc
697	6	2.0	488	2	AF1794	IMP dehydrogenase	770	2.0	522	2	A50478	neuroilin - goldfis
698	6	2.0	489	2	B24177	keratin, 55K type	771	2.0	524	2	A23518	keratin, 57K type
699	6	2.0	489	2	A47259	corticosteroid-bin	772	2.0	524	2	S76140	hypothetical prote
700	6	2.0	490	2	T23791	hypothetical prote	773	2.0	525	2	F84933	IMP cyclohydrolase
701	6	2.0	491	2	S74473	probable starch sy	774	2.0	525	2	A36316	acu-8 protein - Ne
702	6	2.0	491	2	C24829	H+-transporting tw	775	2.0	528	2	S63009	hypothetical prote
703	6	2.0	492	2	S71245	glucose-6-phosphat	776	2.0	528	2	D97270	ATP-dependent RNA
704	6	2.0	493	2	F84689	probable cytochrom	777	2.0	529	1	YRHU1	monophenol monooxy
705	6	2.0	493	2	G84689	probable cytochrom	778	2.0	530	2	H82405	cytochrome d ubiqu
706	6	2.0	493	2	H64072	ribose transport p	779	2.0	530	2	A89783	oligopeptide trans
707	6	2.0	494	2	AH2676	sulfate adenylyate	780	2.0	532	2	T08205	nucleocapsid prote
708	6	2.0	494	2	F97458	hypothetical prote	781	2.0	534	2	A29776	glucan 1,4-alpha-g
709	6	2.0	494	2	G89995	sucrose-6-phosphat	782	2.0	534	2	I37942	keratin 4, type II
710	6	2.0	494	2	S27696	tcnM protein - Str	783	2.0	536	2	C82096	aminoacyl-histidin
711	6	2.0	495	2	H90397	conserved hypothet	784	2.0	536	2	H85647	hypothetical prote
712	6	2.0	496	2	C83122	probable aldehyde	785	2.0	536	2	T27668	hypothetical prote
713	6	2.0	497	2	T48367	hypothetical prote	786	2.0	536	2	T08241	gas-vesicle operon
714	6	2.0	497	2	T51195	hypothetical prote	787	2.0	536	2	S15183	gas-vesicle operon
715	6	2.0	498	2	H81782	adhesin MafB2 NMA2	788	2.0	538	2	T31774	hypothetical prote
716	6	2.0	499	2	T10443	probable major pro	789	2.0	539	2	S57972	hypothetical prote
717	6	2.0	499	2	S70113	hypothetical prote	790	2.0	540	2	B45665	adult-specific 61.
718	6	2.0	500	2	D64889	probable phenylace	791	2.0	543	2	S56830	probable purine nu
719	6	2.0	500	2	T19525	hypothetical prote	792	2.0	544	2	F97170	uncharacterized co
720	6	2.0	500	2	B83910	hypothetical prote	793	2.0	545	1	SYECTP	CIP synthase (EC 6
721	6	2.0	501	2	B26304	ribose transport p	794	2.0	545	2	H91083	CIP synthase lim
722	6	2.0	501	2	AE0952	high affinity ribo	795	2.0	545	2	E82074	CIP synthase VC244
723	6	2.0	501	2	C91215	hypothetical prote	796	2.0	545	2	A85929	CIP synthase (im
724	6	2.0	501	2	D86061	hypothetical prote	797	2.0	545	2	AC0410	CIP synthase (EC 6
725	6	2.0	502	2	A23547	keratin, type II c	798	2.0	545	2	AD0859	CIP synthase (im
726	6	2.0	502	2	D65057	hypothetical prote	799	2.0	547	2	H73211	hypothetical prote
727	6	2.0	502	2	T00483	hypothetical prote	800	2.0	548	2	E71961	probable transport
728	6	2.0	502	2	T25669	hypothetical prote	801	2.0	548	2	T41092	hypothetical prote
729	6	2.0	503	1	CTBPRH	site-specific DNA-	802	2.0	548	2	G70610	hypothetical prote
730	6	2.0	503	2	QJ1019	site-specific DNA-	803	2.0	551	2	B88949	protein R0985_4 li
731	6	2.0	503	2	S25094	keratin, type II,	804	2.0	551	2	S86635	probable membrane
732	6	2.0	504	1	PWTA	H+-transporting tw	805	2.0	553	2	I59009	epidermal keratin
733	6	2.0	504	2	D86332	hypothetical prote	806	2.0	555	2	H89957	formyltetrahydrof
734	6	2.0	505	2	C84949	NADH2 dehydrogenas	807	2.0	556	2	A87185	amidophosphoribosy
735	6	2.0	506	1	PWEGA	H+-transporting tw	808	2.0	556	2	AH1981	hypothetical prote
736	6	2.0	506	2	D81410	hypothetical prote	809	2.0	557	2	D96654	hypothetical prote
737	6	2.0	507	1	PWRZA	H+-transporting tw	810	2.0	557	4	S42226	hypothetical large
738	6	2.0	507	1	PWZMA	H+-transporting tw	811	2.0	558	2	C96598	hypothetical prote
739	6	2.0	507	2	A48661	probable legumin A	812	2.0	558	2	JQ1798	BAR protein - vacc
740	6	2.0	507	2	T06452	cystathionine beta	813	2.0	558	2	G83049	DNA repair protein
741	6	2.0	507	2	T20293	hypothetical prote	814	2.0	558	2	F64402	vanadate-sensitive
742	6	2.0	507	2	T36370	probable sensory h	815	2.0	560	2	S46096	probable membrane
743	6	2.0	508	2	B91250	hypothetical prote	816	2.0	561	2	D70034	oligo-1,6-glucosid
744	6	2.0	508	2	D81417	probable Arp / Grp	817	2.0	561	2	AG2336	potassium-dependen
745	6	2.0	508	2	B96658	hypothetical prote	818	2.0	562	2	T15396	hypothetical prote
746	6	2.0	509	2	T16846	hypothetical prote	819	2.0	564	1	KRHUEB	keratin 6a, type I
747	6	2.0	510	2	S53970	amidophosphoribosy	820	2.0	564	1	KRHUEB	keratin 6b, type I
748	6	2.0	510	2	H90787	Rtn-like protein [	821	2.0	564	2	I61758	keratin 6c, type I
749	6	2.0	510	2	JC4208	nitrogenase NifB c	822	2.0	564	2	I61770	keratin 6e, type I
750	6	2.0	510	2	T09237	NifB protein - Fra	823	2.0	564	2	I61771	keratin 6f, type I
751	6	2.0	511	1	S54720	glucose-6-phosphat	824	2.0	565	2	A83361	penicillin-binding
752	6	2.0	512	2	S41292	glucose-1-phosphat	825	2.0	565	2	JB0160	sodium bicarbonate
753	6	2.0	512	2	S75887	hypothetical prote	826	2.0	567	2	S58750	NADH2 dehydrogenas
754	6	2.0	512	2	C91268	transcription acti	827	2.0	567	2	T40057	probable single-st
755	6	2.0	512	2	G90399	amino acid transpo	828	2.0	569	2	A69511	C conserved hypoth
756	6	2.0	512	2	A86109	transcription acti	829	2.0	571	2	E96550	hypothetical prote
757	6	2.0	512	2	C41968	transcription acti	830	2.0	572	2	S18732	autoantigen, 64K -
758	6	2.0	513	2	S08381	keratin, 58K type	831	2.0	577	2	H82544	pilus biogenesis p
759	6	2.0	514	2	H95050	peptide chain rele	832	2.0	578	2	T35264	probable BCCT fami

833	6	2.0	578	2	B37852	phosphotransferase	906	640	2	C81108	ABC transporter, A
834	6	2.0	579	2	S50446	VAC8 protein - yea	907	641	1	Z22RNQ	adenyl-sulfate k
835	6	2.0	579	2	F69157	exonuclease ABC c	908	641	2	E95320	adenyl-sulfate k
836	6	2.0	581	1	KRMS2	keratin, type II c	909	645	2	A44861	keratin, 67K type
837	6	2.0	582	2	H97017	ATP-dependent Zn p	910	645	2	S54123	neurofilament prot
838	6	2.0	582	2	C59099	hypothetical prote	911	648	2	T04949	hypothetical prote
839	6	2.0	583	2	G35214	ABC transporter, A	912	649	2	A49512	alpha-amylase (EC
840	6	2.0	583	2	S29961	Ref(2)Pp protein -	913	651	2	T50289	WD repeat protein
841	6	2.0	583	2	F98078	neopullulanase (EC	914	653	2	F70383	organic solvent to
842	6	2.0	584	2	B86835	hypothetical prote	915	654	2	F83260	hypothetical prote
843	6	2.0	585	2	A1271	pyruvate kinases h	916	654	2	S69673	SAC7 protein - yea
844	6	2.0	585	2	AD1633	pyruvate kinases h	917	655	1	ALXBS	cyclomalodextrin
845	6	2.0	585	2	T19100	hypothetical prote	918	655	2	T32472	hypothetical prote
846	6	2.0	587	2	F30044	hypothetical prote	919	656	2	S49941	probable membrane
847	6	2.0	587	2	AG3019	conserved hypotet	920	656	2	G85731	Rhs element associ
848	6	2.0	588	2	A71661	hypothetical prote	921	659	2	H81431	methyl-accepting c
849	6	2.0	589	2	B29514	hypothetical prote	922	661	2	AD1989	hypothetical prote
850	6	2.0	589	2	A29476	muscarinic acetyl c	923	662	2	G88451	protein K10D2.6 li
851	6	2.0	589	2	T38718	muscarinic acetyl c	924	668	2	JQ2356	capsid protein - f
852	6	2.0	590	2	T42202	major facilitator	925	669	2	D70369	ATP-dependent DNA
853	6	2.0	590	2	T42206	probable acyl-CoA	926	670	2	T51275	hypothetical prote
854	6	2.0	590	2	A29904	keratin 5, type II	927	671	1	C69621	fructose-bisphosph
855	6	2.0	590	2	S10128	muscarinic acetyl c	928	671	1	QRBSCH	two-component sens
856	6	2.0	590	2	S01114	muscarinic acetyl c	929	673	2	T21007	hypothetical prote
857	6	2.0	590	2	S47572	muscarinic acetyl c	930	674	2	S46092	probable membrane
858	6	2.0	590	2	S29964	ref(2)Pn protein -	931	675	2	T40680	hypothetical prote
859	6	2.0	595	2	T41042	hypothetical prote	932	676	1	S35704	protein kinase C (
860	6	2.0	596	2	A85438	hypothetical prote	933	676	2	T47637	hypothetical prote
861	6	2.0	599	2	S06785	gene ref(2)P prote	934	676	2	T47526	protein kinase-lik
862	6	2.0	599	2	S29963	Ref(2)P02 protein	935	680	2	T39858	probable ABC trans
863	6	2.0	600	2	S65788	ku antigen 70K cha	936	681	2	G90075	hypothetical prote
864	6	2.0	601	2	AF1192	heat shock protein	937	682	2	S68970	triacylglycerol li
865	6	2.0	602	2	C64564	GTP-binding membra	938	682	2	S68970	triacylglycerol li
866	6	2.0	602	2	AC2313	hypothetical prote	939	682	2	F90603	viapl-like (mycopla
867	6	2.0	604	2	A71947	GTP-binding protei	940	691	2	D96805	probable acyl-CoA
868	6	2.0	604	2	F69802	ABC transporter (A	941	692	2	T28783	hypothetical prote
869	6	2.0	605	2	S48940	hypothetical prote	942	694	2	A83126	probable TonB-depe
870	6	2.0	606	2	S25266	hypothetical prote	943	695	2	T28782	hypothetical prote
871	6	2.0	607	2	E70165	hypothetical prote	944	696	2	G97626	hypothetical prote
872	6	2.0	607	2	A35391	hetA protein - Ana	945	696	2	A12849	GGDEF family prote
873	6	2.0	607	2	AD2160	heterocyst differe	946	697	2	T16908	hypothetical prote
874	6	2.0	607	2	H20590	hypothetical prote	947	700	2	F84601	probable ubiquitin
875	6	2.0	612	2	J01346	glucan 1,4-alpha-g	948	702	2	B64250	cell division prot
876	6	2.0	612	2	S73611	glucose inhibited	949	702	2	T39483	probable transmemb
877	6	2.0	614	1	B70772	probable adenyl-l	950	703	2	B82355	hypothetical prote
878	6	2.0	614	2	F86719	hypothetical prote	951	706	1	A45416	protein kinase C (
879	6	2.0	615	2	F87599	hypothetical prote	952	706	2	A86350	F8K7.10 protein -
880	6	2.0	615	2	S64746	ORC3 protein - yea	953	707	1	A44500	protein kinase C (
881	6	2.0	617	2	T16189	hypothetical prote	954	708	2	D81712	type III secretion
882	6	2.0	617	2	AF1284	5-methyltetrahydro	955	709	2	S73497	cell division prot
883	6	2.0	617	2	A11655	5-methyltetrahydro	956	709	2	T28712	hypothetical prote
884	6	2.0	620	1	I39755	adenyl-sulfate k	957	710	1	QOB822	membrane antigen g
885	6	2.0	620	2	AG1598	internalin like pr	958	710	2	T13458	hypothetical prote
886	6	2.0	622	2	T37257	hypothetical prote	959	711	2	AD0901	polynucleotide pho
887	6	2.0	623	2	A48123	cell cycle regulat	960	714	2	AF2458	hypothetical prote
888	6	2.0	627	2	T02415	probable homeodoma	961	716	2	G82627	hypothetical prote
889	6	2.0	628	2	B87448	conserved hypotet	962	719	2	E88504	SecA-related prote
890	6	2.0	628	2	T24234	hypothetical prote	963	729	2	G82437	glycyl-tRNA synthe
891	6	2.0	629	2	T05425	polyadenylate-bind	964	730	2	D86422	protein T4O12.13 [
892	6	2.0	629	2	S42629	keratin K3 - rabbi	965	731	2	A96788	hypothetical prote
893	6	2.0	630	2	B4584	transposition regu	966	734	2	T28744	hypothetical prote
894	6	2.0	632	2	T15954	hypothetical prote	967	734	1	WZBEA1	DNA-packaging prot
895	6	2.0	633	2	H95932	probable adenyl-l	968	734	2	T42588	polyribonucleotide
896	6	2.0	633	2	T14612	hypothetical prote	969	734	2	H65106	polynucleotide pho
897	6	2.0	634	2	B98265	hypothetical prote	970	734	2	E91134	polynucleotide pho
898	6	2.0	635	2	S30018	allantoin transpor	971	734	2	H85979	polynucleotide pho
899	6	2.0	636	2	T35042	probable transcrip	972	735	2	F98228	1,4-alpha-glucan b
900	6	2.0	638	2	I53169	cytokeatin 2 - hu	973	735	2	AH3057	glycogen branching
901	6	2.0	639	1	B41328	1,4-alpha-glucan b	974	735	2	T39464	hypothetical prote
902	6	2.0	639	2	C70154	DNA topoisomerase	975	737	2	S72442	actin-fragmin kina
903	6	2.0	639	2	D70720	hypothetical prote	976	741	2	A26572	hsg25p protein - f
904	6	2.0	640	1	ALASR	glucan 1,4-alpha-g	977	746	2	AD1622	probable integral
905	6	2.0	640	2	A29166	glucan 1,4-alpha-g	978	748	2	S24134	endopeptidase 2 (E

979 6 2.0 748 2 G83927 alpha-galactosidase  
 980 6 2.0 754 2 G83025 copolysomerase IV s  
 981 6 2.0 756 2 F70815 hypothetical prote  
 982 6 2.0 758 1 S45477 SEC18 protein - ye  
 983 6 2.0 759 2 B90892 probable oxidoredu  
 984 6 2.0 759 2 G85725 probable oxidoredu  
 985 6 2.0 759 2 H64903 hypothetical prote  
 986 6 2.0 760 2 C87029 hypothetical prote  
 987 6 2.0 762 2 T01444 proteinase homolog  
 988 6 2.0 763 2 E96571 hypothetical prote  
 989 6 2.0 764 2 I48882 thyrotropin recept  
 990 6 2.0 764 2 A35956 thyrotropin recept  
 991 6 2.0 768 2 G82506 hypothetical prote  
 992 6 2.0 772 2 A48822 protein-glutamine  
 993 6 2.0 773 2 H96818 hypothetical prote  
 994 6 2.0 774 2 S55113 hypothetical prote  
 995 6 2.0 776 2 AH1147 preprotein translo  
 996 6 2.0 776 2 AH1506 preprotein translo  
 997 6 2.0 776 2 I48317 cyclin F - mouse  
 998 6 2.0 780 2 T27941 hypothetical prote  
 999 6 2.0 781 2 AH3355 outer membrane pro  
 1000 6 2.0 786 2 A55501 cyclin F - human

## ALIGNMENTS

RESULT 1  
 G81235  
 50S ribosomal protein L10 NMB0130 [imported] - Neisseria meningitidis (strain MC58 serog  
 C:Species: Neisseria meningitidis  
 C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 02-Feb-2001  
 C:Accession: G81235; F82007  
 R:Testelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A  
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
 ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.  
 Science 287, 1809-1815, 2000  
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
 A>Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
 A:Reference number: AB1000; MUID:20175755; PMID:10710307  
 A:Accession: G81235  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-166 <TET>  
 A:Cross-references: GB:AE002371; GB:AE002098; NID:g7225337; PIDN:AAF40589.1; PID:g722534  
 A:Experimental source: serogroup B, strain MC58  
 R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
 ; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
 Nature 404, 502-506, 2000  
 A>Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
 A:Reference number: AB1775; MUID:20222556; PMID:10761919  
 A:Accession: F82007  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-166 <PAR>  
 A:Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83459.1; PID:g737891  
 A:Experimental source: serogroup A, strain Z2491  
 C:Genetics:  
 A:Gene: rplJ; NMB0130; NMA0144  
 C:Superfamily: Escherichia coli ribosomal protein L10

Query Match 2.7%; Score 8; DB 2; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 3.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 262 VAEIASIP 269  
 |||||  
 Db 122 VAEIASIP 129

RESULT 2  
 T36274  
 probable epimerase - Streptomyces coelicolor

C:Species: Streptomyces coelicolor  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C:Accession: T36274  
 R:Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrrell, B.G.; Rajandream, M.A.  
 Submitted to the EMBL Data Library, June 1999  
 A:Reference number: Z21576  
 A:Accession: T36274  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-353 <MUR>  
 A:Cross-references: EMBL:AL079345; PIDN:CAB45360.1; GSPDB:GN000070; SCOEDB:SCE68.24c  
 A:Experimental source: strain A3(2)  
 C:Genetics:  
 A:Gene: SCOEDB:SCE68.24c

Query Match 2.7%; Score 8; DB 2; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 6.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 SVKETNVI 63  
 |||||  
 Db 87 SVKETNVI 94

## RESULT 3

AF2174  
 hypothetical protein alr2949 [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp.  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C:Accession: AF2174  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Saamamoto, S.; Watanabe, A.; Iriuchi  
 Nakazaki, N.; Shampo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001  
 A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AF2174  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-620 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BAB74648.1; PID:g17132043; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: alr2949

Query Match 2.7%; Score 8; DB 2; Length 620;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 190 YDFSINNL 197  
 |||||  
 Db 558 YDFSINNL 565

## RESULT 4

G83329  
 probable RND efflux transporter PA2526 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C:Species: Pseudomonas aeruginosa  
 C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: G83329  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A:Reference number: AB2950; MUID:20437337; PMID:10984043  
 A:Accession: G83329  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1036 <STO>  
 A:Cross-references: GB:AE004680; GB:AE004091; NID:g9948580; PIDN:AAG05914.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:

A:Gene: PA2526  
C:Superfamily: hypothetical protein b2075

Query Match 2.7%; Score 8; DB 2; Length 1036;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 191 DFLSNLNS 198  
|||||  
Db 383 DFLSNLNS 390

RESULT 5  
T01508  
mismatch repair enzyme T10M13.8 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 24-Mar-1999  
C:Accession: T01508  
R:Johnson, A.F.; de la Bastide, M.; Lodhi, M.; Hoffman, J.; Hasegawa, A.; Gnoj, L.; Goto Martienssen, R.; McCombie, W.  
submitted to the EMBL Data Library, May 1997  
A:Description: The sequence of the Arabidopsis thaliana T10M13 BAC.  
A:Reference number: Z14346  
A:Accession: T01508  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1324 <JOH>  
A:Cross-references: EMBL:AF001308; NID:G2104523; PID:G3912921  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 4S  
A:Introns: 316/1; 349/2; 383/3; 401/3; 426/3; 449/3; 477/3; 542/3; 719/2; 762/3; 793/2;  
A:Note: T10M13.8.

Query Match 2.7%; Score 8; DB 2; Length 1324;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 SLPNISSS 26  
|||||  
Db 855 SLPNISSS 862

RESULT 6  
T23016  
hypothetical protein F59F5.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T23016  
R:Lennard, N.  
submitted to the EMBL Data Library, August 1995  
A:Reference number: Z19653  
A:Accession: T23016  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-93 <WIL>  
A:Cross-references: EMBL:Z50794; PIDN:CAA90658.1; GSPDB:GN00028; CESP:F59F5.4  
A:Experimental source: clone F59F5  
C:Genetics:  
A:Gene: CESP:F59F5.4  
A:Map position: X  
A:Introns: 17/1; 53/3

Query Match 2.4%; Score 7; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 179 LSIISVI 185  
|||||  
Db 19 LSIISVI 25

RESULT 7

A:Gene: PA2526  
C:Superfamily: hypothetical protein b2075

Query Match 2.4%; Score 7; DB 2; Length 146;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 VLVCLLS 19  
|||||  
Db 5 VLVCLLS 11

RESULT 8  
S58350  
vatP protein - slime mold (Dictyostelium discoideum)  
C:Species: Dictyostelium discoideum  
C:Date: 14-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 22-Jun-1999  
C:Accession: S58350  
R:Coukell, M.  
submitted to the EMBL Data Library, August 1995  
A:Reference number: S58350  
A:Accession: S58350  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-196 <COU>  
A:Cross-references: EMBL:X90516; NID:G944998; PIDN:CAA62102.1; PID:G944999  
C:Superfamily: vacuolar H<sup>+</sup>-transporting ATPase 16K chain  
C:Keywords: ATP

Query Match 2.4%; Score 7; DB 2; Length 196;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 83 ASVGISN 89  
|||||  
Db 54 ASVGISN 60

RESULT 9  
D69396  
hypothetical protein AF1173 - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 17-Mar-2000  
C:Accession: D69396  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

A:Gene: PA2526  
C:Superfamily: hypothetical protein b2075

Query Match 2.7%; Score 8; DB 2; Length 1036;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 191 DFLSNLNS 198  
|||||  
Db 383 DFLSNLNS 390

RESULT 5  
T01508  
mismatch repair enzyme T10M13.8 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 24-Mar-1999  
C:Accession: T01508  
R:Johnson, A.F.; de la Bastide, M.; Lodhi, M.; Hoffman, J.; Hasegawa, A.; Gnoj, L.; Goto Martienssen, R.; McCombie, W.  
submitted to the EMBL Data Library, May 1997  
A:Description: The sequence of the Arabidopsis thaliana T10M13 BAC.  
A:Reference number: Z14346  
A:Accession: T01508  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1324 <JOH>  
A:Cross-references: EMBL:AF001308; NID:G2104523; PID:G3912921  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 4S  
A:Introns: 316/1; 349/2; 383/3; 401/3; 426/3; 449/3; 477/3; 542/3; 719/2; 762/3; 793/2;  
A:Note: T10M13.8.

Query Match 2.7%; Score 8; DB 2; Length 1324;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 SLPNISSS 26  
|||||  
Db 855 SLPNISSS 862

RESULT 6  
T23016  
hypothetical protein F59F5.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T23016  
R:Lennard, N.  
submitted to the EMBL Data Library, August 1995  
A:Reference number: Z19653  
A:Accession: T23016  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-93 <WIL>  
A:Cross-references: EMBL:Z50794; PIDN:CAA90658.1; GSPDB:GN00028; CESP:F59F5.4  
A:Experimental source: clone F59F5  
C:Genetics:  
A:Gene: CESP:F59F5.4  
A:Map position: X  
A:Introns: 17/1; 53/3

Query Match 2.4%; Score 7; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 179 LSIISVI 185  
|||||  
Db 19 LSIISVI 25

RESULT 7

A:Gene: PA2526  
C:Superfamily: hypothetical protein b2075

Query Match 2.4%; Score 7; DB 2; Length 146;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 VLVCLLS 19  
|||||  
Db 5 VLVCLLS 11

RESULT 8  
S58350  
vatP protein - slime mold (Dictyostelium discoideum)  
C:Species: Dictyostelium discoideum  
C:Date: 14-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 22-Jun-1999  
C:Accession: S58350  
R:Coukell, M.  
submitted to the EMBL Data Library, August 1995  
A:Reference number: S58350  
A:Accession: S58350  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-196 <COU>  
A:Cross-references: EMBL:X90516; NID:G944998; PIDN:CAA62102.1; PID:G944999  
C:Superfamily: vacuolar H<sup>+</sup>-transporting ATPase 16K chain  
C:Keywords: ATP

Query Match 2.4%; Score 7; DB 2; Length 196;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 83 ASVGISN 89  
|||||  
Db 54 ASVGISN 60

RESULT 9  
D69396  
hypothetical protein AF1173 - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 17-Mar-2000  
C:Accession: D69396  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

B95418  
hypothetical protein SMA2323 [imported] - Sinorhizobium meliloti (strain 1021) megaplasm  
C:Species: Sinorhizobium meliloti  
C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C:Accession: B95418  
R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows  
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.  
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001  
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti  
A:Reference number: A95262; MUID:21396509; PMID:11481432  
A:Accession: B95418  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-146 <KUR>  
A:Cross-references: GB:AE006469; PIDN:AAK65908.1; PID:G14524420; GSPDB:GN00165  
A:Experimental source: strain 1021, megaplasmid pSymA  
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A:Reference number: A96039; MUID:21368234; PMID:11474104  
A:Contents: annotation  
C:Genetics:  
A:Gene: SMA2323  
A:Genome: plasmid

Query Match 2.4%; Score 7; DB 2; Length 146;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 VLVCLLS 19  
|||||  
Db 5 VLVCLLS 11

RESULT 8  
S58350  
vatP protein - slime mold (Dictyostelium discoideum)  
C:Species: Dictyostelium discoideum  
C:Date: 14-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 22-Jun-1999  
C:Accession: S58350  
R:Coukell, M.  
submitted to the EMBL Data Library, August 1995  
A:Reference number: S58350  
A:Accession: S58350  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-196 <COU>  
A:Cross-references: EMBL:X90516; NID:G944998; PIDN:CAA62102.1; PID:G944999  
C:Superfamily: vacuolar H<sup>+</sup>-transporting ATPase 16K chain  
C:Keywords: ATP

Query Match 2.4%; Score 7; DB 2; Length 196;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 83 ASVGISN 89  
|||||  
Db 54 ASVGISN 60

RESULT 9  
D69396  
hypothetical protein AF1173 - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 17-Mar-2000  
C:Accession: D69396  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997  
A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiaich, P.; Kaine, B.P.; Sykes, S.  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon  
A:Reference number: A69250; MUID:98049343; PMID:9389475  
A:Accession: D69396  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-197 <KLE>  
A:Cross-references: GB:AE001023; GB:AE000782; NID:G2689346; PIDN:AAB90085.1; PID:G264942  
C:Superfamily: Archaeoglobus hypothetical protein Af2020

Query Match 2.4%; Score 7; DB 2; Length 197;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 136 PGGYTLS 142  
|||||  
Db 175 PGGYTLS 181

RESULT 10  
B83142  
hypothetical protein PA4038 [imported] - Pseudomonas aeruginosa (strain PA01).  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
A:Accession: B83142  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: B83142  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-244 <STO>  
A:Cross-references: GB:AE004820; GB:AE004091; NID:G9950223; PIDN:AAG07425.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA4038

Query Match 2.4%; Score 7; DB 2; Length 244;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 TVLVCLL 18  
|||||  
Db 167 TVLVCLL 173

RESULT 11  
T04642  
hypothetical protein F10N7.160 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 20-Sep-1999  
A:Accession: T04642  
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hohseil, J.; Mewes, H.W.; Mayer, K.F.X  
submitted to the Protein Sequence Database, March 1999  
A:Reference number: Z15263  
A:Accession: T04642  
A:Molecule type: DNA  
A:Residues: 1-253 <BEV>  
A:Cross-references: EMBL:AL021636  
A:Experimental source: cultivar Columbia; BAC clone F10N7  
C:Genetics:  
A:Map position: 4  
A:Introns: 128/3; 147/3; 168/3; 191/3  
A:Note: F10N7.160  
C:Superfamily: Arabidopsis hypothetical protein F10N7.160

Query Match 2.4%; Score 7; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 49;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 70 LKQDVDS 76  
|||||  
Db 75 LKQDVDS 81

RESULT 12  
G70635  
hypothetical protein Rv193lc - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 15-Sep-2000  
A:Accession: G70635  
R:Cole, S.T.; Broch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: G70635  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-259 <COL>  
A:Cross-references: GB:284498; GB:AL123456; NID:G3261701; PIDN:CAB06530.1; PID:e293526;  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: Rv193lc  
C:Superfamily: Mycobacterium tuberculosis hypothetical protein Rv193lc

Query Match 2.4%; Score 7; DB 2; Length 259;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 205 GGAGVDA 211  
|||||  
Db 18 GGAGVDA 24

RESULT 13  
D64782  
3-hydroxyisobutyrate dehydrogenase homolog b0509 - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
A:Accession: D64782  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co.  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: D64782  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-292 <BLAT>  
A:Cross-references: GB:AS000157; GB:U00096; NID:G1786716; PIDN:AAC73611.1; PID:G1786719;  
A:Experimental source: strain K-12, substrain MGL655  
C:Superfamily: 3-hydroxyisobutyrate dehydrogenase; 3-hydroxyisobutyrate dehydrogenase ho  
F13-263/Domain: 3-hydroxyisobutyrate dehydrogenase homology <HIB>

Query Match 2.4%; Score 7; DB 2; Length 292;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 284 GGEIGAR 290  
|||||  
Db 121 GGEIGAR 127

RESULT 14  
B90700  
translational semialdehyde reductase [imported] - Escherichia coli (strain O157:H7, substra  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 24-Aug-2001



C;Accession: B90700  
 R;Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic islands  
 A;Reference number: A99629; MUID:21156231; PMID:11258796  
 A;Accession: B90700  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-292 <HAY>  
 A;Cross-references: GB:BA000007; PIDN:BA033993.1; PID:g13360028; GSPDB:GN00154  
 A;Experimental source: strain O157:H7, substrain RMD 050952  
 C;Genetics:  
 A;Gene: EC0570  
 C;Superfamily: 3-hydroxyisobutyrate dehydrogenase; 3-hydroxyisobutyrate dehydrogenase h0

Query Match 2.4%; Score 7; DB 2; Length 292;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 284 GGEIGAR 290  
 |||||  
 Db 121 GGEIGAR 127

RESULT 15  
 B85550  
 probable oxidoreductase ybbQ [imported] - *Escherichia coli* (strain O157:H7, substrain ED3)  
 C;Species: *Escherichia coli*  
 C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C;Accession: B85550  
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimallanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
 A;Reference number: A85480; MUID:21074935; PMID:11206551  
 A;Accession: B85550  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-292 <STO>  
 A;Cross-references: GB:AB005174; NID:g12513403; PIDN:AAG54865.1; GSPDB:GN00145; UWGP:Z06  
 A;Experimental source: strain O157:H7, substrain EDL933  
 C;Genetics:  
 A;Gene: ybbQ  
 C;Superfamily: 3-hydroxyisobutyrate dehydrogenase; 3-hydroxyisobutyrate dehydrogenase h0

Query Match 2.4%; Score 7; DB 2; Length 292;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 284 GGEIGAR 290  
 |||||  
 Db 121 GGEIGAR 127

RESULT 16  
 AB0567  
 2-hydroxy-3-oxopropionate reductase (EC 1.1.1.60) [imported] - *Salmonella enterica* subsp.  
 C;Species: *Salmonella enterica* subsp. *enterica* serovar Typhi  
 A;Note: this species has also been called *Salmonella typhi*  
 C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
 C;Accession: AB0567  
 R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
 th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
 S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A;Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov  
 A;Reference number: AB0502; PMID:11677608  
 A;Accession: AB0567  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-292 <PAR>

A;Cross-references: GB:AL513382; PIDN:CAD05004.1; PID:g16501788; GSPDB:GN00176  
 C;Genetics:  
 A;Gene: garR  
 C;Superfamily: 3-hydroxyisobutyrate dehydrogenase; 3-hydroxyisobutyrate dehydrogenase h0  
 C;Keywords: oxidoreductase

Query Match 2.4%; Score 7; DB 2; Length 292;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 284 GGEIGAR 290  
 |||||  
 Db 121 GGEIGAR 127

RESULT 17  
 AD3495  
 DNA-formamidopyrimidine glycosylase (EC 3.2.2.23) [imported] - *Brucella melitensis* (stra  
 C;Species: *Brucella melitensis*  
 C;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 03-Jun-2002  
 C;Accession: AD3495  
 R;DelVecchio, V.G.; Kaputal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,  
 Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier,  
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 A;Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*  
 A;Reference number: AD3252; PMID:11756688  
 A;Accession: AD3495  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-293 <KUR>  
 A;Cross-references: GB:AE008917; PIDN:AAL53127.1; PID:g17983995; GSPDB:GN00190  
 A;Experimental source: strain 16M  
 C;Genetics:  
 A;Gene: BMEI1946  
 A;Map position: I  
 C;Superfamily: formamidopyrimidine-DNA glycosylase  
 C;Keywords: glycosylase; hydrolase

Query Match 2.4%; Score 7; DB 2; Length 293;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 177 DGLSIIS 183  
 |||||  
 Db 66 DGLSIIS 72

RESULT 18  
 H81395  
 probable ferroxidase (EC 4.99.1.1) Cj0503c [imported] - *Campylobacter jejuni* (strain  
 C;Species: *Campylobacter jejuni*  
 C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
 C;Accession: H81395  
 R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell  
 Nature 403, 665-668, 2000  
 A;Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyp  
 A;Reference number: A81250; MUID:20150912; PMID:10688204  
 A;Accession: H81395  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-303 <PAR>  
 A;Cross-references: GB:AL1139075; GB:AL111168; NID:g6967817; PIDN:CAB75140.1; PID:g696797  
 A;Experimental source: serotype O2, strain NCTC 11168  
 C;Genetics:  
 A;Gene: Cj0501; Cj0503c  
 C;Keywords: lyase

Query Match 2.4%; Score 7; DB 2; Length 303;  
 Best Local Similarity 100.0%; Pred. No. 58;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 AYQSKLG 229

Db 223 AYQSKLG 229  
|||||  
RESULT 19  
S58090  
probable membrane protein YDR090c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein Y06652.02c  
C:Species: Saccharomyces cerevisiae  
C>Date: 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 19-Apr-2002  
C:Accession: S58090  
R:Oliver, K.; Harris, D.  
submitted to the EMBL Data Library, July 1995  
A:Reference number: S58089  
A:Accession: S58090  
A:Molecule type: DNA  
A:Residues: 1-310 <OLI>  
A:Cross-references: EMBL:Z50111; NID:g914872; PID:g914874; GSPDB:GN000004; MIPS:YDR090c  
A:Experimental source: strain AB972  
C:Genetics:  
A:Gene: MIPS:YDR090c  
A:Cross-references: SGD:S0002497  
A:Map position: 4R  
C:Keywords: transmembrane protein  
F:7-23/Domain: transmembrane #status predicted <TM1>  
F:43-59/Domain: transmembrane #status predicted <TM2>  
F:97-113/Domain: transmembrane #status predicted <TM3>  
F:132-148/Domain: transmembrane #status predicted <TM4>  
F:168-184/Domain: transmembrane #status predicted <TM5>  
F:189-205/Domain: transmembrane #status predicted <TM6>  
Query Match 2.4%; Score 7; DB 2; Length 310;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 179 LSIIISVI 185  
|||||  
Db 177 LSIIISVI 183  
RESULT 20  
AH0634  
probable 2-hydroxyacid dehydrogenase in phoB-csg intergenic region STY1172 [imported]  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: this species has also been called Salmonella typhi  
C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AH0634  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A:Reference number: AB0502; PMID:11677608  
A:Accession: AH0634  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-312 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD08259.1; PID:gi6502306; GSPDB:GN00176  
C:Genetics:  
A:Gene: STY1172  
C:Superfamily: phosphoglycerate dehydrogenase  
Query Match 2.4%; Score 7; DB 2; Length 312;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 206 GAGVDAT 212  
|||||  
Db 64 GAGVDAT 70  
RESULT 21

AH2476  
permease protein of sugar ABC transporter alr5368 [imported] - Nostoc sp. (strain PCC 71  
C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AH2476  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AH2476  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-312 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BAB7067.1; PID:gl7134507; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr5368  
C:Superfamily: probable ribose ABC transporter rbsC-2  
Query Match 2.4%; Score 7; DB 2; Length 312;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 272 TSAVATL 278  
|||||  
Db 51 TSAVATL 57  
RESULT 22  
D97349  
glycosyltransferase [imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C>Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C:Accession: D97349  
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4836, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: D97349  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-322 <KUR>  
A:Cross-references: GB:AF001437; PIDN:AAK81583.1; PID:gl5026763; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC3661  
C:Superfamily: stress response protein csbB  
Query Match 2.4%; Score 7; DB 2; Length 322;  
Best Local Similarity 100.0%; Pred. No. 61;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 62 VITKNLI 68  
|||||  
Db 251 VITKNLI 257  
RESULT 23  
TI4615  
hypothetical protein - Trypanosoma cruzi  
C:Species: Trypanosoma cruzi  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: TI4615  
R:Andersson, B.; Aslund, L.; Pettersson, U.  
submitted to the EMBL Data Library, March 1998  
A:Description: 93.4 kb of complete sequence from chromosome 3 of Trypanosoma cruzi.  
A:Reference number: Z18159  
A:Accession: TI4615  
A:Status: preliminary  
A:Molecule type: DNA

A;Residues: 1-334 <AND>  
A;Cross-references: EMBL:AF052832; NID:g3063540; PID:g3063552; PIDN:AAC14088.1  
C;Genetics:  
A;Map position: 3

Query Match 2.4%; Score 7; DB 2; Length 334;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 286 EIGARLT 292  
|||||||  
Db 195 EIGARLT 201

RESULT 24  
F86787  
metal ABC transporter substrate binding protein yndG [imported] - Lactococcus lactis subsp. lactis  
C;Species: Lactococcus lactis subsp. lactis  
C;Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C;Accession: F86787  
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich, S.; Artigues, C.; et al.  
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis subsp. lactis  
A;Reference number: A86625; MUID:21235186; PMID:11337471  
A;Accession: F86787  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-334 <STO>  
A;Cross-references: GB:AE005176; PID:g12724280; PIDN:AAK05400.1; GSPDB:GN00146  
A;Experimental source: strain IL1403  
C;Genetics:  
A;Gene: yndG

Query Match 2.4%; Score 7; DB 2; Length 334;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 66 NLIALKK 72  
|||||||  
Db 131 NLIALKK 137

RESULT 25  
T31269  
4-hydroxy-2-oxovalerate aldolase (EC 4.1.3.-) - Sphingomonas aromaticivorans plasmid pNL1  
C;Species: Sphingomonas aromaticivorans  
C;Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 11-Jan-2000  
C;Accession: T31269  
R;Romine, M.F.; Stillwell, L.C.; Wong, K.K.; Thurston, S.J.; Sisk, E.C.; Sensen, C.W.; et al.  
A;Description: Complete sequence of a 184 kb catabolic plasmid from Sphingomonas aromaticivorans  
A;Reference number: Z20992  
A;Accession: T31269  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-343 <ROW>  
A;Cross-references: EMBL:AF079317; NID:g3378261; PID:g3378410; PIDN:AAD03993.1  
C;Genetics:  
A;Gene: xylK  
A;Keywords: plasmid pNL1

Query Match 2.4%; Score 7; DB 2; Length 343;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 207 AGVDAIE 213  
|||||||  
Db 43 AGVDAIE 49

RESULT 26  
T25622  
hypothetical protein C37H5.1 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000  
C;Accession: T25622  
R;Davidson, S.; Gillam, B.  
A;Description: The sequence of C. elegans cosmid C37H5.  
A;Reference number: Z20058  
A;Accession: T25622  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-351 <DAV>  
A;Cross-references: EMBL:U89315; PIDN:AAB42365.1; GSPDB:GN00023; CESP:C37H5.1  
A;Experimental source: strain Bristol N2; clone C37H5  
C;Genetics:  
A;Gene: CESP:C37H5.1  
A;Map position: 5  
A;Introns: 42/1; 73/3; 130/3; 154/3; 257/1; 294/3  
C;Superfamily: annexin I; annexin repeat homology

Query Match 2.4%; Score 7; DB 2; Length 351;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 TIINTVL 14  
|||||||  
Db 47 TIINTVL 53

RESULT 27  
C69348  
hypothetical protein AF0787 - Archaeoglobus fulgidus  
C;Species: Archaeoglobus fulgidus  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
C;Accession: C69348  
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Nature 390, 364-370, 1997  
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artigues, C.; Venter, J.C.; Smith, H.O.; Woese, C.R.; et al.  
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
A;Reference number: A69250; MUID:98049343; PMID:9389475  
A;Accession: C69348  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-358 <KLE>  
A;Cross-references: GB:AE001050; GB:AE000782; NID:g2689373; PIDN:AAB90458.1; PID:g264982

Query Match 2.4%; Score 7; DB 2; Length 358;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 180 SIISVIV 186  
|||||||  
Db 128 SIISVIV 134

RESULT 28  
T24223  
hypothetical protein R13H4.3 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
C;Accession: T24223  
R;Kishaw, J.  
A;Description: The sequence of R13H4.3  
A;Reference number: Z19858  
A;Accession: T24223  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-376 <WIL>  
A;Cross-references: EMBL:Z81579; PIDN:CAB04655.1; GSPDB:GN00023; CESP:R13H4.3  
A;Experimental source: clone R13H4

C:Genetics:  
A:Gene: CESP:R13H4.3  
A:Map position: 5  
A:Introns: 59/3; 90/3; 144/3; 284/3; 324/3  
C:Superfamily: mammalian acid phosphatase

Query Match 2.4%; Score 7; DB 2; Length 376;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 SNFSVKE 59  
|||||||  
DB 84 SNFSVKE 90

RESULT 29  
B82973  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: B82973  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B. adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, ; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: B82973  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-384 <STO>  
A:Cross-references: GB:AE004091; NID:99951705; PIDN:AAG08775.1; GSPDB:GN001178  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA5390

Query Match 2.4%; Score 7; DB 2; Length 384;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 GEIGARL 291  
|||||||  
DB 210 GEIGARL 216

RESULT 30  
C83102  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: C83102  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B. adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, ; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: C83102  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-388 <STO>  
A:Cross-references: GB:AE004851; GB:AE004091; NID:99950571; PIDN:AAG07743.1; GSPDB:GN001178  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA4355  
C:Superfamily: Streptomyces lividans chloramphenicol resistance protein

Query Match 2.4%; Score 7; DB 2; Length 388;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 275 VATLNIG 281  
|||||||

DB 324 VATLNIG 330

RESULT 31  
AF1739  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AF1739  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H. D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A:Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AF1739  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-433 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CAC97686.1; PID:gl6414981; GSPDB:GN001178  
A:Experimental source: strain Clp11262  
C:Genetics:  
A:Gene: lin2459

Query Match 2.4%; Score 7; DB 2; Length 433;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NKJLFTI 9  
|||||||  
DB 156 NKJLFTI 162

RESULT 32  
H97237  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: H97237  
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, ; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: H97237  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-455 <KUB>  
A:Cross-references: GB:AE001437; PIDN:AAK80691.1; PID:gl5025782; GSPDB:GN001168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC2745

Query Match 2.4%; Score 7; DB 2; Length 455;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 PSNISLF 241  
|||||||  
DB 139 PSNISLF 145

RESULT 33  
AG1368  
C:Species: Mus musculus (house mouse)  
C:Date: 09-Sep-1994 #sequence\_revision 09-Sep-1994 #text\_change 07-May-1999  
C:Accession: AG1368  
R:Yu, D.; Pang, Y.Y.; Checkla, D.M.; Freedberg, I.M.; Sun, T.T.; Bertolino, A.P. J. Invest. Dermatol. 97, 354-363, 1991  
A:Title: Transient expression of mouse hair keratins in transfected HeLa cells: interact

A:Reference number: A61368; MUID:91302895; PMID:1712823

A:Accession: A61368

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-479 <YUA>

C:Superfamily: cytoskeletal keratin

Query Match 2.4%; Score 7; DB 2; Length 479;  
Best Local Similarity 100.0%; Pred. No. 88;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 ALKKDVD 75

Db 208 ALKKDVD 214

#### RESULT 34

S05408

keratin, type II, component 7c, cytoskeletal - sheep (tentative sequence)

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C>Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 31-Mar-2000

C:Accession: S05408

R:Sparrow, L.G.; Robinson, C.P.; McMahon, D.T.W.; Rubira, M.R.

Biochem. J. 261, 1015-1022, 1989

A:Title: The amino acid sequence of component 7c, a type II intermediate-filament protein

A:Reference number: S05408; MUID:90026244; PMID:2803231

A:Accession: S05408

A:Molecule type: protein

A:Residues: 1-491 <SPA>

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil

Query Match 2.4%; Score 7; DB 2; Length 491;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 ALKKDVD 75

Db 213 ALKKDVD 219

#### RESULT 35

E69487

protein translocase, subunit SEC61 alpha (secY) homolog - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Sep-1999

C:Accession: E69487

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes, S.

Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaea

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: E69487

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-493 <KLE>

A:Cross-references: GB:A6000971; GB:A6000782; NID:92689294; PIDN:AAB89347.1; PID:9264863

C:Superfamily: yeast SSH1 protein

Query Match 2.4%; Score 7; DB 2; Length 493;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 118 EGTRVEI 124

Db 263 EGTRVEI 269

#### RESULT 36

I57463

keratin type II - mouse (fragment)

C:Species: Mus musculus (house mouse)

C>Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 13-Aug-1999

C:Accession: I57463

R:Tobiasch, E.; Schweizer, J.; Winter, H.

Mol. Biol. Rep. 16, 39-47, 1992

A:Title: Structure and site of expression of a murine type II hair keratin.

A:Reference number: I57463; MUID:92186831; PMID:1372089

A:Accession: I57463

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-495 <RES>

A:Cross-references: GB:M92088; NID:9198283; PIDN:AAA39273.1; PID:9198284

C:Superfamily: cytoskeletal keratin

Query Match 2.4%; Score 7; DB 2; Length 495;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 ALKKDVD 75

Db 224 ALKKDVD 230

#### RESULT 37

T34712

probable hydrolase - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C>Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999

C:Accession: T34712

R:Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, May 1998

A:Reference number: Z21554

A:Accession: T34712

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-506 <OLI>

A:Cross-references: EMBL:AL023702; PIDN:CAA19244.1; GSPDB:GN00070; SCOEDB:SC1C3.20c

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC1C3.20c

Query Match 2.4%; Score 7; DB 2; Length 506;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 206 GAGVDAT 212

Db 281 GAGVDAT 287

#### RESULT 38

KRSHL2

keratin type II, microfibrillar - sheep

N:Alternate names: intermediate filament protein KII-9; low-sulfur keratin

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C>Date: 15-Nov-1984 #sequence\_revision 15-May-1998 #text\_change 10-Dec-1999

C:Accession: I46409; A02954; S22025

R:Powell, B.; Crocker, L.; Rogers, G.

Development 114, 417-433, 1992

A:Title: Hair follicle differentiation: expression, structure and evolutionary conservat

A:Reference number: I46409; MUID:92274852; PMID:1375545

A:Accession: I46409

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-508 <PO2>

A:Cross-references: EMBL:X62509; NID:g1307; PIDN:CAA44368.1; PID:g1308

R:Crewther, W.G.; Inglis, A.S.; McKern, N.M.

Biochem. J. 173, 365-371, 1978

A:Title: Amino acid sequences of alpha-helical segments from S-carboxymethylkerateine-A.

A:Reference number: A02954; MUID:7902075; PMID:581264

A:Accession: A02954

A:Molecule type: protein

A;Residues: 142-190,192-251 <CRE>  
A;Note: 51-Lys and 52-Lys were also found  
C;Genetics:  
A;Gene: KII-9  
A;Introns: 123/3; 193/3; 214/3; 246/3; 301/3; 343/3; 417/2; 428/1  
C;Superfamily: cytoskeletal keratin  
C;Keywords: coiled coil; cytoskeleton; hair; intermediate filament  
F;126-418/Domain: rod #status predicted <ROD>  
F;126-253/Region: coiled coil 1 #status predicted  
F;271-418/Region: coiled coil 2 #status predicted

Query Match 2.4%; Score 7; DB 1; Length 508;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 ALKKDQVD 75  
|||||  
DB 211 ALKKDQVD 217

RESULT 39  
F72619  
probable NADH dehydrogenase (ubiquinone) chain 13 APE1415 - Aeropyrum pernix (strain K1)  
C;Species: Aeropyrum pernix  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C;Accession: F72619  
R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999  
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix  
A;Reference number: A72450; MUID:99310339; PMID:10382966  
A;Accession: F72619  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-513 <KAW>  
A;Cross-references: DDBJ:AP000061; NID:g5104821; PIDN:BAA80412.1; PID:g5105098  
A;Experimental source: strain K1  
C;Genetics:  
A;Gene: APE1415  
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 4

Query Match 2.4%; Score 7; DB 2; Length 513;  
Best Local Similarity 100.0%; Pred. No. 94;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 LGIAYSL 234  
|||||  
DB 136 LGIAYSL 142

RESULT 40  
S33561  
ref(2)P protein - fruit fly (Drosophila erecta)  
C;Species: Drosophila erecta  
C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 21-Jul-2000  
C;Accession: S33561; S29962  
R;Dru, P.; Bras, F.; Dezelee, S.; Gay, P.; Petitjean, A.M.; Pierre-Deneubourg, A.; Tenin Genetics 133, 943-954, 1993  
A;Title: Unusual variability of the Drosophila melanogaster ref(2)P protein which controls the expression of the ref(2)P gene  
A;Reference number: S33561; MUID:93216091; PMID:8462852  
A;Accession: S33561  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-594 <DRU>  
A;Cross-references: EMBL:X69831; NID:g7407; PIDN:CAA49485.1; PID:g7408  
A;Note: submitted to the EMBL Data Library, December 1992  
C;Genetics:  
A;Gene: FlyBase:Dere/ref(2)P  
A;Cross-references: FlyBase:FBgn0012278  
A;Introns: 58/1; 547/1

Query Match 2.4%; Score 7; DB 2; Length 594;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 89 NPSNFTI 95  
|||||  
DB 113 NPSNFTI 119

Search completed: July 8, 2003, 10:33:50  
Job time : 69 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	840	100.0	840	24	ABK68876		DNA encoding Ehrli
2	600.2	71.5	852	20	AAK34744		DNA encoding Omp-1
3	450.8	53.7	495	20	AAK34770		DNA encoding P30-1
4	238.4	28.4	924	20	AAK34761		DNA encoding P30-1
5	238.4	28.4	1607	21	AAO1292		Ehrlichia canis im
6	238.4	28.4	1607	24	ABK68852		DNA encoding Ehrli
7	234	27.9	840	21	AAO1294		Ehrlichia canis im
8	234	27.9	840	21	ABK68854		DNA encoding Ehrli
9	234	27.9	2037	21	AAO1295		Ehrlichia canis im

XX Claim 5; Figure 14; 106pp; English.

XX The invention relates to a recombinant homologous 28 kDa immunodominant protein, p28 (I) of Ehrlichia canis. (I), a 28-kDa antigen preferably dispersed in a pharmaceutically acceptable carrier, is useful for inhibiting E. canis infection in a subject. (I) is useful in the development of vaccines and serodiagnostics that are particularly effective for disease prevention and serodiagnosis. ABK68852-ABK68878 represent the 28-kDa antigen coding sequences and PCR primers of the invention.

XX Sequence 840 BP; 301 A; 141 C; 144 G; 254 T; 0 other;

Query Match 100.0%; Score 840; DB 24; Length 840;  
Best Local Similarity 100.0%; Pred. No. 1.1e-188;  
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAATTAAGAAATTTCTAGTAAGAGCGGTTAATCTCAATTAATGTCATCTTACCA 60  
Db 1 ATGAATTAAGAAATTTCTAGTAAGAGCGGTTAATCTCAATTAATGTCATCTTACCA 60

QY 61 TATCAGTCTTTGACAGCTCTGAGGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120  
Db 61 TATCAGTCTTTGACAGCTCTGAGGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120

QY 121 ATTAGTGAAGTACAATCAAGTATATACACTTTAGAAATTTCTCTGCTGAAGAACT 180  
Db 121 ATTAGTGAAGTACAATCAAGTATATACACTTTAGAAATTTCTCTGCTGAAGAACT 180

QY 181 CCTATTATGAACAAATTTCTCTCACTAAAGATTTTCGGACTAAAGAGAGTGTGAT 240  
Db 181 CCTATTATGAACAAATTTCTCTCACTAAAGATTTTCGGACTAAAGAGAGTGTGAT 240

QY 241 ATACAAAAAGACGATTTTACAAGAGTAGTCCAGGCAATGATTTTCAAAATACCTTA 300  
Db 241 ATACAAAAAGACGATTTTACAAGAGTAGTCCAGGCAATGATTTTCAAAATACCTTA 300

QY 301 ATATCAGATTTTCAGGAGTATTTGGTACTCTATGCGAGGACCAAGAAATAGAACTTGA 360  
Db 301 ATATCAGATTTTCAGGAGTATTTGGTACTCTATGCGAGGACCAAGAAATAGAACTTGA 360

QY 361 GCTGCATATCACAATTTAATCCAAAAACACCGATACAAATGATCTGATAATGCTGAA 420  
Db 361 GCTGCATATCACAATTTAATCCAAAAACACCGATACAAATGATCTGATAATGCTGAA 420

QY 421 TACTATAAATTTTGCATTTATCTCGTAAAGATGCAATGGAAGATCAGCAATATAGTA 480  
Db 421 TACTATAAATTTTGCATTTATCTCGTAAAGATGCAATGGAAGATCAGCAATATAGTA 480

QY 481 CTTAAAAATGACGGCAATACCTTTTATGTCATTTGATGGTTAATACCTGCTATGACATTA 540  
Db 481 CTTAAAAATGACGGCAATACCTTTTATGTCATTTGATGGTTAATACCTGCTATGACATTA 540

QY 541 GCTGAAGAGATCTTTTGCATACCATATGCTGAGGTATAGGAGCAGATCTTATCACT 600  
Db 541 GCTGAAGAGATCTTTTGCATACCATATGCTGAGGTATAGGAGCAGATCTTATCACT 600

QY 601 ATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGAAAAATAGGTATTAGTTACCT 660  
Db 601 ATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGAAAAATAGGTATTAGTTACCT 660

QY 661 ATCACACGAGATCTGTCATTTATTTGGTGGATACCTACCATGGGTTATTGGTAAATAA 720  
Db 661 ATCACACGAGATCTGTCATTTATTTGGTGGATACCTACCATGGGTTATTGGTAAATAA 720

QY 721 TTTGAGAAGATACCTGTATTAATCTCTGTAGTATTAATGATGCTCCACACCATCT 780  
Db 721 TTTGAGAAGATACCTGTATTAATCTCTGTAGTATTAATGATGCTCCACACCATCT 780

QY 781 GCTTCAGTAACCTTTGACGTTTGGATATCTTTGGCGGAGAAATTTGGAATGAGGTTTCACTTC 840  
Db 781 GCTTCAGTAACCTTTGACGTTTGGATATCTTTGGCGGAGAAATTTGGAATGAGGTTTCACTTC 840

RESULT 2  
AAAX34744  
ID AAX34744 standard; DNA; 852 BP.  
XX  
AC AAX34744;  
XX  
DT 05-JUL-1999 (first entry)  
XX  
DE DNA encoding OMP-1B protein.  
XX  
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
XX detection; dog; ss.  
XX  
OS Ehrlichia chaffeensis.  
XX  
PN WO9913720-A1.  
XX  
PD 25-MAR-1999.  
XX  
PF 18-SEP-1998; 98WO-US19600.  
XX  
PR 19-SEP-1997; 97US-0059353.  
XX  
PA (OHIS ) UNIV OHIO STATE.  
XX  
PI Ohashi N, Rikihisa Y;  
XX  
DR WPI; 1999-254290/21.  
XX  
DR P-PSDB; AAY06944.  
XX  
PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
PT Ehrlichia canis  
XX  
PS Disclosure; Fig 4A; 55pp; English.  
XX  
CC The invention provides isolated outer membrane proteins (OMP) from  
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
CC in AAY06943-958. The E. canis proteins form part of the P30 family and  
CC consist of proteins shown in AAY06959-970. The proteins and genes are  
CC used to detect E. chaffeensis in patients and E. canis in dogs.  
XX  
SQ Sequence 852 BP; 301 A; 155 C; 157 G; 239 T; 0 other;

Query Match 71.5%; Score 600.2; DB 20; Length 852;  
Best Local Similarity 82.7%; Pred. No. 4e-132;  
Matches 702; Conservative 0; Mismatches 138; Indels 9; Gaps 1;

QY 1 ATGAATTAAGAAATTTCTAGTAAGAGCGGTTAATCTCAATTAATGTCATCTTACCA 60  
Db 1 ATGAATTAAGAAATTTCTAGTAAGAGCGGTTAATCTCAATTAATGTCATCTTACCT 60

QY 61 TATCAGTCTTTGACAGATCTGTAAGTTCATCTCACTAAAAAGTTTTCGGCTGAAGAA 111  
Db 61 TATCAGTCTTTGACAGATCTGTAAGTTCATCTCACTAAAAAGTTTTCGGCTGAAGAA 111

QY 61 TACCAATCTTTGACAGATCTGTAAGTTCATCTCACTAAAAAGTTTTCGGCTGAAGAA 120  
Db 61 TACCAATCTTTGACAGATCTGTAAGTTCATCTCACTAAAAAGTTTTCGGCTGAAGAA 120

QY 112 GGCTTCTACATTTAGTCAAGTACAACTCAAGTATATCACATTTAGAAAAATTTCTGCT 171  
Db 112 GGCTTCTACATTTAGTCAAGTATATCACTCAAGTATATCACATTTAGAAAAATTTCTGCT 171

QY 121 GGCTTCTACATTTAGTCAAGTATATCACTCAAGTATATCACATTTAGAAAAATTTCTGCT 180  
Db 121 GGCTTCTACATTTAGTCAAGTATATCACTCAAGTATATCACATTTAGAAAAATTTCTGCT 180

QY 172 GAAGAAATCTCTATTATTAAGAAATTTCTCACTAAAAAGTTTTCGGCTGAAGAA 231  
Db 172 GAAGAAATCTCTATTATTAAGAAATTTCTCACTAAAAAGTTTTCGGCTGAAGAA 231

QY 181 GAAGAAATCTCTATTATTAAGAAATTTCTCACTAAAAAGTTTTCGGCTGAAGAA 240  
Db 181 GAAGAAATCTCTATTATTAAGAAATTTCTCACTAAAAAGTTTTCGGCTGAAGAA 240

QY 232 GATGCTGATATAACAAAAAGACGATTTTACAGAGTAGCTCCAGGCAATTTGATTTTCAA 291  
Db 232 GATGCTGATATAACAAAAAGACGATTTTACAGAGTAGCTCCAGGCAATTTGATTTTCAA 291

QY 241 GACGAGATATAGCAATCTGCGAATTTTACAGGACAGATCCAGCTCGATTTTCAAG 300  
Db 241 GACGAGATATAGCAATCTGCGAATTTTACAGGACAGATCCAGCTCGATTTTCAAG 300

QY 292 AATAACTTAATATCAGGATTTTCAGGAGTATTTGTTTCTCTATGAGCGGACCAAGATA 351  
Db 292 AATAACTTAATATCAGGATTTTCAGGAGTATTTGTTTCTCTATGAGCGGACCAAGATA 351

QY 301 AATAACTTAATATCAGGATTTTCAGGAGTATTTGTTTCTCTATGAGCGGACCAAGATA 360  
Db 301 AATAACTTAATATCAGGATTTTCAGGAGTATTTGTTTCTCTATGAGCGGACCAAGATA 360



```

Db 594 TAGTATCCATGTTGAAGCTACAAATCTTAAATTTCTTACCAAGGAAGTAGGTTTA 653
Qy 652 AGTTACCCCTATCACACAGAGAGTCTCTGCATTTATTTGGTGGATACCATCGGCGTTATT 711
Db 654 AGCTACTCTATAAGCCAGAGCTCTGTGTTTATTTGGTGGCATTTCATAAGGTAATA 713
Qy 712 GGTATAAATTTGAGAAGATAC---CTGTATAACTCTCTGTAGTATTAAATGATGCTCCT 768
Db 714 GGAACGAAATTTAGAGATATTTCTTACTATAATCTTACTGATCAACACTTGCAGGAAA 773
Qy 769 CAACACCATCTGTTCTAGTACTCTTGACCTTGGATCTTGGCGGAGAAATTGGA 825
Db 774 GGAACCTACCTGCAATAGTAATACTGATGTATGCCACTTTGGAATAGAAATGGGA 830

RESULT 13
AAV07179
-ID AAV07179 standard; DNA; 4683 BP.
XX AC AAV07179;
XX DT 14-SEP-1998 (first entry)
XX DE Ehrlichia chaffeensis VSA genomic locus.
XX KW MAP1 homologue; variable surface antigen; VSA1; VSA2; VSA3; VSA4;
XX KW VSA5; rickettsia; DNA vaccine; ss.
XX OS Ehrlichia chaffeensis.
XX FH Key Location/Qualifiers
XX CDS 3..134
XX FT /*tag= a
XX FT /note= "VSA1 partial gene"
XX FT terminator 203..212
XX FT /*tag= b
XX FT terminator 226..239
XX FT /*tag= c
XX FT -35_signal 349..354
XX FT /*tag= d
XX FT -10_signal 375..380
XX FT /*tag= e
XX FT RBS 430..434
XX FT /*tag= f
XX FT CDS 439..1299
XX FT /*tag= g
XX FT /note= "VSA2"
XX FT terminator 1300..1309
XX FT /*tag= h
XX FT terminator 1324..1333
XX FT /*tag= i
XX FT misc_feature 1347..1361
XX FT /*tag= j
XX FT /note= "G-rich region"
XX FT -35_signal 1473..1478
XX FT /*tag= k
XX FT -10_signal 1499..1504
XX FT /*tag= l
XX FT RBS 1554..1558
XX FT /*tag= m
XX FT CDS 1563..2399
XX FT /*tag= n
XX FT /note= "VSA3"
XX FT terminator 2444..2456
XX FT /*tag= o
XX FT terminator 2465..2477
XX FT /*tag= p
XX FT misc_feature 2491..2505
XX FT /*tag= q
XX FT /note= "G-rich region"
XX FT -35_signal 2618..2623
XX FT /*tag= r

```

```

FT -10_signal 2644..2649
FT /*tag= b
FT RBS 2699..2703
FT /*tag= t
FT CDS 2708..3550
FT /*tag= u
FT /note= "VSA4"
FT terminator 3586..3596
FT /*tag= v
FT terminator 3610..3622
FT /*tag= w
FT misc_feature 3586..3596
FT /*tag= x
FT /note= "G-rich region"
FT -35_signal 3767..3772
FT /*tag= y
FT -10_signal 3788..3793
FT /*tag= z
FT RBS 3847..3851
FT /*tag= aa
FT CDS 3856..4683
FT /*tag= ab
FT /note= "truncated VSA5 gene"
XX W09816554-A1.
PN 23-APR-1998.
XX 17-OCT-1997; 97WO-US19044.
XX 17-OCT-1996; 96US-0733230.
XX (UYFL ) UNIV FLORIDA.
XX Barbet AF, BurrIDGE MJ, Ganta RR, Mahan SM, McGuire TC;
XX Nyika A, Rurangirwa FR;
XX WPI; 1998-251232/22.
XX P-PSDB; AAW51091-05.
XX Composition containing nucleic acid encoding rickettsial antigen -
XX useful for, e.g. stimulating protective immune response in humans or
XX animals
XX Claim 4; Fig 2A-B; 39pp; English.
XX This is the DNA sequence of a 4.6 kb genomic locus of Ehrlichia
XX chaffeensis that was obtained using a PCR cloning strategy based on
XX identifying genes homologous to the major antigenic protein MAPI
XX (see AAW51088) of Cowdria ruminantium. It includes 5 very similar
XX but non-identical open reading frames (ORFs), of which ORF1 is a
XX partial gene and ORF5 is nearly complete but lacks 5-7 amino acid
XX codons and a termination codon (see AAW51091-95). Due to their
XX similarity to MAPI surface antigen genes of C. ruminantium, the
XX E. chaffeensis ORFs are designated variable surface antigen (VSA)
XX genes 1-5. A claimed composition comprises a nucleic acid (see
XX AAV07176-82) encoding a polypeptide (see AAW51088-99) that elicits a
XX protective immune response against a rickettsial pathogen. The
XX nucleic acid is used, in human or veterinary medicine, in vaccines
XX to protect against Rickettsia, Ehrlichia, Anaplasma and Cowdria
XX species. The nucleic acids are also useful as probes to identify
XX related sequences, e.g. for identification of organisms and for
XX diagnosing infection. Use of nucleic acid vaccines avoids the
XX problem of protein purification associated with protein-based
XX vaccines. The nucleic acid does not replicate in the host but
XX remains episomal and capable of expressing polypeptide for at least
XX 19 mth.
XX Sequence 4683 BP; 1590 A; 753 C; 818 G; 1522 T; 0 other;

```

Query Match 27.2%; Score 228.6; DB 19; Length 4683;  
 Best Local Similarity 58.2%; Pred. No. 2.6e-44;  
 Matches 487; Conservative 0; Mismatches 329; Indels 21; Gaps 4;

```
QY 1 ATGAATTATAGAAAATTTCTAGTAGAAGCGGTTAAATCTATTAATGTCAATCTTACCA 60
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3856 ATGAATTATAGAAAATTTCTAGTAGAAGCGGTTAAATCTATTAATGTCAATCTTACCT 3915
QY 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3916 GGAGTATCAATTTCCGACCCAGCAGTAGTGTATTAACGGTAAT-----TCTTAC 3966
QY 121 ATTAGTCAAAAGTACAATCCAAAGTATATCACATTTTAGAAAATTTCTCTGCT---GAAGAA 177
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3967 ATCAGTGGAAAATACATGCCAAGTCTTCGCATTTTGGAGTATTCTCTGCTAAGCAAGAA 4026
QY 178 ACTCCTATTATGGAAACAATTTCTCTCACTAAAAAAGTTTTCGGACTAAAGAAAGATGGT 237
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4027 AGAAATACAACTGTTGGAGTGTTCGACTGAAGCAAAAATTTGGACGGACGCAATATCC 4086
QY 238 GATATAACAAAAAGACGATTTTACAAGAGTAGCTCCAGGCAATTTGATTTTCAAAATAAC 297
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4087 AACTCCTCCCAACGATGATTTCTACTCTCAAAATTTATTCATTTAATATGAAAACAC 4146
QY 298 TTAATATCAGGATTTTTCAGGAAGTATTGGTTACTTCTATGGACCGACCAAGAAATAGACTT 357
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4147 CCGTTTTTAGGTTTTTCAGGAGCTATTGGTTACTCAATGGATGGTCCCAAGAAATAGAGCTT 4206
QY 358 GAGCTGCATATCAACAATTTAATCCAAAACACCGATAACAATGATGATAATGGT 417
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4207 GAAGTATCTTATGAACAATTTGATGTAAAAATCAAGGTAAACAATTAAGAAATGAAGCA 4266
QY 418 GAATACTATAAATTTTGCATTATCTCTGTAAGATGCAATG-----GAAGATCAGCAA 471
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4267 CATAGATATTGCTCTATCCCACTCACTCAGCAGCAGATGAGTAGTGCAAGTAATAT 4326
QY 472 TAGTAGTACTTAAAAATGACCGCATAACTTTTATGTCAATGATGGTTAATCTTCTAT 531
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4327 TTGTCTTTCTAAAAATGAAGGATTAATCTTGACATATCATTTATGCTGAACGATGCTAT 4386
QY 532 GACATTACAGCTGAAGGATATCTTTCGTACCATATGATGTCAGGTATAGGACAGAT 591
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4387 GAGCTAGTAGGCGAGGATACCTTTTCTCTTATATGCGCAGGTATCGGTACTGAT 4446
QY 592 CTATCACTATTTTAAAGACCTCAATCTAAAATTTGCTTACCAAGGAAAAATAGGTATT 651
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4447 TTAGTATCCATGTTTGAAGCTACAATCTTAAAATTTCTTACCAAGGAAAAGTTAGTTTA 4506
QY 652 AGTTACCTTATCACACCAAGTCTCTGCATTTATTTGGTGGATATCAATGCGCTTAT 711
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4507 AGCTACTCTATAAGCCCAAGGCTTCTGTGTTTATTTGGTGGCACTTTTATAAGGTAATA 4566
QY 712 GGTATAAATTTGAGAGATAC---CTGTAATACTCTGTAGTATTAATGATGCTCT 768
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4567 GGAACGAAATTTAGAGATATTCCTACTATAATACCTTCTGATCAACACTTCGAGAA 4626
QY 769 CAACACCATCTGCTTTCAGTACTCTTGACGTTGGATATCTTTGGCGGAGAAATTTGGA 825
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4627 GGAACCTACCTGCAATAGTATATCTGGATGATGCCACTTTGGATAGAAATGGA 4683
```

## RESULT 14

AAC68716

ID AAC68716 standard; DNA; 4683 BP.

XX AAC68716;

AC AAC68716;

XX 02-MAR-2001 (first entry)

DT Ehrlichia chaffeensis 28 kDa gene locus.

DE Ehrlichia chaffeensis; MAP1; major antigenic protein 1; antirickettsial;

XX Ehrlichia chaffeensis; MAP1; major antigenic protein 1; antirickettsial;

KW vaccine; gene therapy; Rickettsia; Cowdria; Anaplasma; map2; ihworf3;

KW 4hworfl; 18hworfl; 3gdorfl; ds.

XX Ehrlichia chaffeensis.

OS Ehrlichia chaffeensis.

```
XX WO200065063-A2.
XX 02-NOV-2000.
XX 21-APR-2000; 2000WO-US10886.
XX 22-APR-1999; 99US-0130725.
XX (UYFL) UNIV FLORIDA.
XX Barbet AF, Bowie MV, Ganta RR, Burrige MJ, Mahan SM, McGuire TC;
XX Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Allen AR;
XX WPI; 2000-679675/66.
XX New polynucleotides useful as DNA vaccines for conferring immunity to
XX rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
XX comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
XX
XX Example 2; Fig 2A-2B; 63pp; English.
XX The present sequence is given in a specification relating to nucleic
XX acid vaccines which may be used to protect animals or humans against
XX rickettsial diseases caused by a organisms of Rickettsia sp.,
XX Ehrlichia sp., Anaplasma sp. or Cowdria sp. The vaccine elicits an
XX immune response protective against the rickettsial pathogen. The
XX vaccine comprises the major antigenic protein 1 (MAP1) or major
XX antigenic protein 2 (MAP2) of rickettsial pathogens. The nucleic acid
XX vaccines can be driven by the human cytomegalovirus (HCMV)
XX enhancer-promoter. Cowdria ruminantium genes designated map 2, ihworf3,
XX 4hworfl, 18hworfl and 3gdorfl may be used in therapeutic and diagnostic
XX applications. The polypeptides are useful for detecting antibodies
XX associated with infection by a rickettsial pathogen whilst the
XX polynucleotides may be used to detect the presence of rickettsial
XX nucleic acids.
XX Sequence 4683 BP; 1576 A; 747 C; 831 G; 1526 T; 3 other;
XX
XX Query Match 27.2%; Score 228.6; DB 21; Length 4683;
XX Best Local Similarity 58.2%; Pred. No. 2.6e-44;
XX Matches 487; Conservative 0; Mismatches 329; Indels 21; Gaps 4;
QY 1 ATGAATTATAGAAAATTTCTAGTAGAAGCGGTTAAATCTATTAATGTCAATCTTACCA 60
DB 3856 ATGAATTATAGAAAATTTCTAGTAGAAGCGGTTAAATCTATTAATGTCAATCTTACCT 3915
QY 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120
DB 3916 GGAGTATCAATTTCCGACCCAGCAGTAGTGTATTAACGGTAAT-----TCTTAC 3966
QY 121 ATTAGTCAAAAGTACAATCCAAAGTATATCACATTTTAGAAAATTTCTCTGCT---GAAGAA 177
DB 3967 ATCAGTGGAAAATACATGCCAAGTCTTCGCATTTTGGAGTATTCTCTGCTAAGCAAGAA 4026
QY 178 ACTCCTATTATGGAAACAATTTCTCTCACTAAAAAAGTTTTCGGACTAAAGAAAGATGGT 237
DB 4027 AGAAATACAACTGTTGGAGTGTTCGACTGAAGCAAAAATTTGGACGGACGCAATATCC 4086
QY 238 GATATAACAAAAAGACGATTTTACAAGAGTAGCTCCAGGCAATTTGATTTTCAAAATAAC 297
DB 4087 AACTCCTCCCAACGATGATTTCTACTCTCAAAATTTATTCATTTAATATGAAAACAC 4146
QY 298 TTAATATCAGGATTTTTCAGGAAGTATTGGTTACTTCTATGGACCGACCAAGAAATAGACTT 357
DB 4147 CCGTTTTTAGGTTTTTCAGGAGCTATTGGTTACTCAATGGATGGTCCCAAGAAATAGAGCTT 4206
QY 358 GAGCTGCATATCAACAATTTAATCCAAAACACCGATAACAATGATGATAATGGT 417
DB 4207 GAAGTATCTTATGAACAATTTGATGTAAAAATCAAGGTAAACAATTAAGAAATGAAGCA 4266
QY 418 GAATACTATAAATTTTGCATTATCTCTGTAAGATGCAATG-----GAAGATCAGCAA 471
DB 4267 CATAGATATTGCTCTATCCCACTCACTCAGCAGCAGATGAGTAGTGCAAGTAATAT 4326
QY 472 TAGTAGTACTTAAAAATGACCGCATAACTTTTATGTCAATGATGGTTAATCTTCTAT 531
DB 4327 TTGTCTTTCTAAAAATGAAGGATTAATCTTGACATATCATTTATGCTGAACGATGCTAT 4386
QY 532 GACATTACAGCTGAAGGATATCTTTCGTACCATATGATGTCAGGTATAGGACAGAT 591
DB 4387 GAGCTAGTAGGCGAGGATACCTTTTCTCTTATATGCGCAGGTATCGGTACTGAT 4446
QY 592 CTATCACTATTTTAAAGACCTCAATCTAAAATTTGCTTACCAAGGAAAAATAGGTATT 651
DB 4447 TTAGTATCCATGTTTGAAGCTACAATCTTAAAATTTCTTACCAAGGAAAAGTTAGTTTA 4506
QY 652 AGTTACCTTATCACACCAAGTCTCTGCATTTATTTGGTGGATATCAATGCGCTTAT 711
DB 4507 AGCTACTCTATAAGCCCAAGGCTTCTGTGTTTATTTGGTGGCACTTTTATAAGGTAATA 4566
QY 712 GGTATAAATTTGAGAGATAC---CTGTAATACTCTGTAGTATTAATGATGCTCT 768
DB 4567 GGAACGAAATTTAGAGATATTCCTACTATAATACCTTCTGATCAACACTTCGAGAA 4626
QY 769 CAACACCATCTGCTTTCAGTACTCTTGACGTTGGATATCTTTGGCGGAGAAATTTGGA 825
DB 4627 GGAACCTACCTGCAATAGTATATCTGGATGATGCCACTTTGGATAGAAATGGA 4683
```

```
Db 4267 CATAGATATTGCTCTATCCATAACTCAGCAGACACATGAGTAGTGCAGTAATAAT 4326
Qy 472 TATGTAGTACTTAAATACGGCATAACTTTTATGTCATTGATGGTTAACTTGCTAT 531
Db 4327 TTGTCTTTCTAAATAATGAGGATTACTTGACATATCATTTATGCTGAACGATGCTAT 4386
Qy 532 GACATTACAGCTGAAGAGTAGTATCTTTCGTACCATATGCATGTCAGGTATAGGAGCAGAT 591
Db 4387 GACGTAGTAGCGAAGGCATACCTTTTCTCTTATATATGCGCAGGTATCGGTACTGAT 4446
Qy 592 CTTATCACTATTTTAAAGACCTCAACTTAAATTTGCTTACCAAGGAAATAGGTATT 651
Db 4447 TTAGTATCCATGTTTGAAGCTACAAATCTTAAATTTCTTACCAAGGAAATAGGTATT 4506
Qy 652 AGTTACCCATCATCACAGAGAGTCTCTGCAATTTATTTGTTGATGATACCATGGCGTTATT 711
Db 4507 AGCTACTCTATAAGCCAGAGACTTCTGTGTTTATTTGGTGGCACTTTTCAAGGTATA 4566
Qy 712 GGTAAATAATTGAGAGATAC---CTGTAATACTCTCTGATGATTTAAATGATGCTCCT 768
Db 4567 GGGAAACGAATTTAGAGATATTTCTTACTATATATACCTACTGGATCAACACTTGCAGGAAA 4636
Qy 769 CAACACATCTGTTTCACTTCACTTTCAGTCACTTTCAGCTTGATGATCTTGGCGGAGAAATTTGA 825
Db 4627 GGAAACTACCTGCAATAGTAATACTGGATGTATGCCACTTTGGAATAGAAATGGGA 4683

RESULT 15
AAS07578
ID AAS07578 standard; DNA; 4683 BP.
XX AC AAS07578;
XX XX
DT 23-OCT-2001 (first entry)
DE DNA encoding variable surface antigens 1-5 (VSA1-5) from B. chaffeensis.
XX KW Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;
XX KW infection; heartwater; diagnostic; variable surface antigen; VSA; ds.
XX OS Ehrlichia chaffeensis.
XX FH Key Location/Qualifiers
FT CDS 3..134
FT FT /*tag= a
FT FT /product= "Variable surface antigen 1 (VSA1)"
FT FT /partial
FT FT /note= "No start codon"
FT FT 179..188
FT FT /*tag= b
FT FT /note= "Transcription terminator of VSA1 gene"
FT FT 202..212 C
FT FT /*tag= C
FT FT /note= "Transcription terminator of VSA1 gene"
FT FT 226..239
FT FT /*tag= d
FT FT /note= "GC-rich sequence of VSA2 gene"
FT FT 349..354
FT FT /*tag= e
FT FT /note= "-35 region of VSA2 gene"
FT FT 375..380
FT FT /*tag= f
FT FT /note= "-10 region of VSA2 gene"
FT FT 430..434
FT FT /*tag= g
FT FT /note= "Ribosome binding site of VSA2 gene"
FT FT 439..1299
FT FT /*tag= h
FT FT /product= "Variable surface antigen 2 (VSA2)"
FT FT 1300..1309
FT FT /*tag= i
FT FT /note= "Transcription termination signal of VSA2"
```

```
FT terminator 1324..1333
FT /*tag= j
FT /note= "Transcription termination signal of VSA2"
FT GC_signal 1349..1361
FT /*tag= k
FT /note= "GC-rich sequence of VSA3 gene"
FT -35_signal 1473..1478
FT /*tag= l
FT /note= "-35 region of VSA3 gene"
FT -10_signal 1499..1504
FT /*tag= m
FT /note= "-10 region of VSA3 gene"
FT RBS 1534..1558
FT /*tag= n
FT /note= "Ribosome binding site of VSA3 gene"
FT CDS 1563..2399
FT /*tag= o
FT /product= "Variable surface antigen 3 (VSA3)"
FT terminator 2443..2455
FT /*tag= p
FT /note= "Transcription termination signal of VSA3"
FT terminator 2465..2477
FT /*tag= q
FT /note= "Transcription termination signal of VSA3"
FT GC_signal 2491..2505
FT /*tag= r
FT /note= "GC-rich sequence of VSA4 gene"
FT -35_signal 2618..2623
FT /*tag= s
FT /note= "-35 region of VSA4 gene"
FT -10_signal 2644..2649
FT /*tag= t
FT /note= "-10 region of VSA4 gene"
FT RBS 2699..2703
FT /*tag= u
FT /note= "Ribosome binding site of VSA4 gene"
FT CDS 2708..3550
FT /*tag= v
FT /product= "Variable surface antigen 4 (VSA4)"
FT terminator 3596..3596
FT /*tag= w
FT /note= "Transcription termination signal of VSA4"
FT terminator 3610..3622
FT /*tag= x
FT /note= "Transcription termination signal of VSA4"
FT GC_signal 2491..2505
FT /*tag= y
FT /note= "GC-rich sequence of VSA5 gene"
FT -35_signal 3767..3772
FT /*tag= z
FT /note= "-35 region of VSA5 gene"
FT -10_signal 3788..3793
FT /*tag= aa
FT /note= "-10 region of VSA5 gene"
FT RBS 3847..3851
FT /*tag= ab
FT /note= "Ribosome binding site of VSA5 gene"
FT CDS 3856..4683
FT /*tag= ac
FT /product= "Variable surface antigen 5 (VSA5)"
FT /partial
FT /note= "No stop codon"

US6251872-B1.
PD 26-JUN-2001.
XX PF 17-OCT-1997; 97US-0953326.
XX PR 17-OCT-1996; 96US-0733230.
XX PA (UYFL ) UNIV FLORIDA.
XX XX
```

PI Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;  
 PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;  
 XX  
 DR WPI; 2001-424487/45.  
 DR P-PSDB; AAU04195, AAU04196, AAU04197, AAU04198, AAU04199.  
 XX  
 PT New MAP2 genes and polypeptides useful as vaccines for conferring  
 PT immunity to human and animal rickettsial diseases, e.g. heartwater, or  
 PT as molecular markers in nucleic acid analysis procedures -  
 XX  
 PS Example 3; Fig 2A-2B; 30pp; English.  
 XX  
 CC The sequence represents the coding sequence of variable surface antigen  
 CC (VSA) gene locus encoding VSA-5 proteins of Ehrlichia chaffeensis, which  
 CC have similarity to major antigen proteins (MAP). The MAP polynucleotides  
 CC and polypeptides are useful as vaccines for conferring immunity to  
 CC rickettsia infection, including Cowdria ruminantium causing heartwater.  
 CC The MAP polynucleotides may be used as molecular markers in nucleic acid  
 CC analysis procedures, and to produce the MAP polypeptides, which may  
 CC be used to raise antibodies that are reactive with the polypeptides.  
 CC The nucleic acids may further be used as probes to identify  
 CC complementary sequences within other nucleic acid molecules or genomes,  
 CC where such probes can be applied to identify or distinguish infectious  
 CC strains of organisms in diagnostic procedures or in rickettsial  
 CC research where identification of particular organisms or strains is  
 CC needed.  
 XX  
 SQ Sequence 4683 BP; 1578 A; 739 C; 833 G; 1533 T; 0 other;

Query Match 27.2%; Score 228.6; DB 22; Length 4683;  
 Best Local Similarity 58.2%; Pred. No. 2.6e-44; Indels 21; Gaps 4;  
 Matches 487; Conservative 0; Mismatches 329;  
 QY 1 ATGAATTATAGAAAATTTCTAGTAAAGACGCGTTAATCTCAATTAATGTCAATCTTACCA 60  
 DB 3856 ATGAATTACAAAAGTTTTCATAACAGTGCATTGATATCATTAATCTCTCTACT 3915  
 QY 61 TATCAGTCTTTTGCAGACTCTGTAGTTTCAAGAACTAATGATAACAAAGAGCTTCTAC 120  
 DB 3916 GGAGTATCATTTTCCGACCCAGCAGGTAGTGTATTAACGGTAAT-----TTCTAC 3966  
 QY 121 ATTAGTGAACAGTACAACTCAAGTATATCACACTTTAGAAAATTTCTTGCT---GAAGAA 177  
 DB 3967 ATCAGTGAATAATACATGCAAGTCTGCGATTTTGGAGTATTTCTTGCTTAAGAGAA 4026  
 QY 178 ACTCCTATTAAATGGAACAAATTTCTCTCAATAAAAAGTTTTCGGACTAAAGAAAGATGGT 237  
 DB 4027 AGAAATACAAACAGTTGGAGTGTGGACTGAAGCAAAATTTGGACGGAAGCGCAATATCC 4086  
 QY 238 GATATAACAAAAGACAGATTTTACAGAGTAGTCTCCAGGCAATGATTTTCAAAATAC 297  
 DB 4087 AACTCTCCCAACAGATGTATTCTACTGTCTCAAAATTTATCAATATGAAAACAAC 4146  
 QY 298 TTAATATCAGGATTTTCAGGAGTATTTGTTACTCTATGCGCGACCAAGATAGAACTT 357  
 DB 4147 CGGTTTTAGGTTTTCAGGAGCTATTTGTTACTCAATGGATGTTCCAGAAATAGAGCTT 4206  
 QY 358 GAAGCTGCATATCAACAAATTTAATCCAAAAACACCGATAACAAATGATATGATAATGGT 417  
 DB 4207 GAAGTATCTTATGAACATTTTGTATGTAAAAAATCAAGGTAAACAATTAAGAATGAAGCA 4266  
 QY 418 GAATACTATAACATTTTGCATATTTCTGTAAGATGCAATG-----GAAGATCAGCA 471  
 DB 4267 CATAGATATTTGCTCTATCCCAATCACTCAGCAGCAGACATGAGTAGTGCAGATTAAT 4326  
 QY 472 TATGTAGTACTTAAAAATGACGGCATACTTTTATGTCTATTGATGGTTAATCTTGCTAT 531  
 DB 4327 TTTGCTTTTCAAAAATGAAGGATTTACTTGACATATCATTTATGCTGNAACCATGCTAT 4386  
 QY 532 GACATTACAGCTGAAGGAGTATCTTTTCGTAACCATATGCAATGCAAGGTATAGGACAGAT 591  
 DB 4387 GACGTAGTAGGGGAGGCACTCTTTTCTCTTATATATGCGCAGGTATCGGTACTGAT 4446

QY 592 CTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTTACCAAGAAAAAATAGGTATT 651  
 DB 4447 TTAGTATCCATGTTTGAAGCTACAAATCTTAAATTTCTTACCAGGAAGATTAGTTTA 4506  
 QY 652 AGTTACCCCTATCACCACAGAGTCTCTGCATTTTATTTGGTGATACCTACCATGGCGTTATT 711  
 DB 4507 AGCTACTCTATAAGCCCAAGAGCTTCTGTGTTTATTTGGTGCACTTTTCATAAGGTAATA 4566  
 QY 712 GGTAAATAATTTGAGAAGATAC----CTGTAATAACTCTCTGTAGTATTAAATGATGCTCCT 768  
 DB 4567 GGGAAACGAATTTAGAGATATTTCTTACTATAATACTACTTGGATCAACACTTTGCAGGAAA 4626  
 QY 769 CAACACCATCTGCTTTCAGTAACTCTTGAGCTTGGATACCTTTGGCGGAGAAATTTGGA 825  
 DB 4627 GGAAACTACCTGCAATAGTAATACTGGATGTATGCCACTTTGGAAATAGAANTTTGGA 4683  
 RESULT 16  
 AAX34759  
 ID AAX34759 standard; DNA; 867 BP.  
 XX  
 AC AAX34759;  
 XX  
 DT 05-JUL-1999 (first entry)  
 XX  
 DE DNA encoding P30 protein.  
 XX  
 KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 KW detection; dog; ss.  
 OS Ehrlichia canis.  
 XX  
 PN WO9913720-A1.  
 XX  
 PD 25-MAR-1999.  
 XX  
 PF 18-SEP-1998; 98WO-US19600.  
 XX  
 PR 19-SEP-1997; 97US-0059353.  
 XX  
 PA (OHIS) UNIV OHIO STATE.  
 XX  
 PI Ohashi N, Rikihisa Y;  
 XX  
 DR WPI; 1999-254290/21.  
 DR P-PSDB; AAY06959.  
 XX  
 PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
 PT Ehrlichia canis  
 XX  
 PS Disclosure; Fig 19A; 55pp; English.  
 XX  
 CC The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and  
 CC consist of proteins shown in AAY06959-970. The proteins and genes are  
 CC used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX  
 SQ Sequence 867 BP; 305 A; 150 C; 157 G; 255 T; 0 other;

Query Match 25.9%; Score 217.4; DB 20; Length 867;  
 Best Local Similarity 57.8%; Pred. No. 7.7e-42;  
 Matches 502; Conservative 0; Mismatches 331; Indels 36; Gaps 5;  
 QY 1 ATGAATTATAGAAAATTTCTAGTAAAGACGCGTTAATCTCAATTAATGTCAATCTTACCA 60  
 DB 1 ATGAATTCGAAAAGTTTTTTCATAGCAAGTGCATTGATATCACTAATGCTTTCTTACCT 60  
 QY 61 TATCAGTCTTTTGCAGATCCTGTAGTTTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120  
 DB 61 AGCGTATCTTTTCTGTAATCAATACATGAAG-----ATAATATAAATGGTAACCTTTTAC 114

QY 121 ATTAGTGCAGAGTACAATCCAGTATATACACATTTAGAAAATTTCTGCTGAAGAAACT 180  
 Db 115 ATTAGTGCAGAGTATATCCAGTGCCTCACAATTTTGGCGTATTTTTCAGTTTAAAGAGAG 174  
 QY 181 CCTATTAAATGGAAAC--AAATTTCTCACTAAAAAGTTTTCGGAC-----TAAAG 228  
 Db 175 AAAACACACACACTGGAGTTTTCGGATTAATAAACAAGATTGGACGGAGCAACATAAAG 234  
 QY 229 AAAGATGGTGATATAACAAAAAGACGATTTTACAAGAGTAGTCTCCAGGCAATTTGATTT- 287  
 Db 235 GATGCAAGCAGCAGCAGCACACAATAGACCAAGTACAATATTTCTCCATTTCAAATTTATTC 294  
 QY 288 -----TCAAATAACTTAATATATACAGATTTTTCAGGAAGTATTTGGTACTCTATGAC 339  
 Db 295 TTTAAATATGAACAATCCATTTTGGGTTTTCAGGAGCTATTTGGCTACTCAATGGGT 354  
 QY 340 GGACCAAGATAGAACTTTGAGCTGCATATCAACAATTTTAAATCCAAAAACACCGATAAC 399  
 Db 355 GGTCCAAAGGTAGAGTTTGAAGTGTCTTACGAAATATTTTGAATGTAATAAACAAGGTAAAC 414  
 QY 400 AATGATCTGATAATGGTGAATACATATAAACAATTTTGCATTTATCTCGTAAAGATGCAATG 459  
 Db 415 AGTTACAAGAGGATGCTCACAATATTTGGCGTTTATCAAGACACACCGGAGGTATGCCA 474  
 QY 460 GAAG-----ATCAGCAATATAGTAGTACTTAAATATGACGGCAATATTTTATGTCA 510  
 Db 475 CAAGCCGGTCAATCAAAATAAATTTGTCTTCTTAAAAAATGAAGGATTTACTTTGACATATCA 534  
 QY 511 TTGATGGTTATACATTTGATGACATTTACAGCTGAAGGAGTATCTTTTCGTACCATATGCA 570  
 Db 535 CTTATGATAACGCAATGTTATGATATAACAATCGACAGCATGCCATTTTCTCCATATATA 594  
 QY 571 TGTGAGGTATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCT 630  
 Db 595 TGTGAGGTATTTGGTGTAGTACTAGTTTGGATTTTGAACCTCAAAATCTTAAATTTCT 654  
 QY 631 TACCAAGAAATAGGTATTTAGTTACCTATCACACAGAGTCTCTGCAATTTATTTGGT 690  
 Db 655 TATCAAGAAATATAGGTGTAGTTTACTTCCATAAGCCCAAGAGCATCTGTTTTTGTGGA 714  
 QY 691 GGATCTACCATGGGTTTATGGTAATAAATTTGAGAGATACCTGTAAATACCTCTGTA 750  
 Db 715 GGACATTTACAGATTTATAGGTATGAAATTTAAGACATTTCTTGCATTAATCTCTGCT 774  
 QY 751 GTATTAATATGATCTCTCAACACACATCTGCTTCAGTAACTCTTTGAGTTGGATCTTT 810  
 Db 775 GGAGCAACAGAAATTAAGGCGACACAGTTTACAACAGTAAACATTAACATATGCCACTTC 834  
 QY 811 GCGGAGAAATTTGGAATGAGTTTCACTTT 839  
 Db 835 GGACTAGAGCTTGGAGGCGAGTTTACTTTT 863

## RESULT 17

AAX34746  
 ID AAX34746 standard; DNA; 861 BP.

AC AAX34746;

DT 05-JUL-1999 (first entry)

XX DNA encoding OMP-1D protein.

DE Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 KW detection; dog; ss.

XX Ehrlichia chaffeensis.

OS WO9913720-A1.

XX 25-MAR-1999.

XX 18-SEP-1998; 98WO-US19600.

XX 19-SEP-1997; 97US-0059353.  
 PR (OHIS ) UNIV OHIO STATE.  
 PA Ohashi N, Rikihisa Y;  
 XX WPI; 1999-254290/21.  
 DR P-PSDB; AAY06946.  
 XX Novel outer membrane proteins from Ehrlichia chaffeensis and  
 PT Ehrlichia canis  
 XX Disclosure; Fig 6A; 55pp; English.

The invention provides isolated outer membrane proteins (OMP) from  
 Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 in AAY06943-958. The E. canis proteins form part of the P30 family and  
 consist of proteins shown in AAY06959-970. The proteins and genes are  
 used to detect E. chaffeensis in patients and E. canis in dogs.

SQ Sequence 861 BP; 281 A; 140 C; 169 G; 271 T; 0 other;

Query Match 25.3%; Score 212.8; DB 20; Length 861;  
 Best Local Similarity 57.3%; Pred. No. 9.4e-41;  
 Matches 492; Conservative 0; Mismatches 342; Indels 24; Gaps 5;

QY 1 ATGAATTTATAAGAAATTTCTAGTAAGAGCGGTTAATCTCATTAATGCTCAATCTTACCA 60  
 Db 1 ATGAACCTGCGAAAAATTTTATTAACAACACTGCATTAACATTAATGCTCTCTTCTTACCT 60  
 QY 61 TATCAGCTCTTTTGCAGATCTGTAGGTTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120  
 Db 61 GGAATATCACTTTCTGATCCAGTACAG-----GATGACAACATAGTGGTAAATTTCTAC 114  
 QY 121 ATTAGTCAAGATACAATCCAGATATATCACATTTTAGAAAAATTTCTGTCT---GAAGAA 177  
 Db 115 ATCAGTGGAAAGTATATGCAAGCGCTTCGCAATTTTGGAGTTTCTGCAAGAGAGAA 174  
 QY 178 ACTCCTATTATGAAACAAATTTCTCTCACTAAAAAGTTTTCGCACTAAAGAAAGTGGT 237  
 Db 175 AGAATACACAGTTGGAGTATTTGGAAATAGACAAGATTTGGGATAGATGTTGTAATATCT 234  
 QY 238 GATATAACAAAAAGACGATTTTACAAGAGTAGTCCAGGCATTTGATTTTCAAAAATAAC 297  
 Db 235 AGAACCACTTTAAGCATATATTCACCGTTCCAAATTTATTCATTTAGTAGTGAATAAT 294  
 QY 298 TTAATATCAGGATTTTTCAGGAAGTATTTGGTTACTCTATGACGCGCAACGAATAGAACTT 357  
 Db 295 CTATTTTTCAGGATTTTGCAGGAGCTATTTGGGCTACTCAATGGATGGCCCAAGAATAGAGCTT 354  
 QY 358 GAAGCTGCATATCAACAATTTTATCCAAAAACACCGCATTAACATGATGATGATAATG-- 415  
 Db 355 GAAGTATCTTATGAAGCATTTGATGTTTAAAAAATCAAGGTAAACAATTTAAGACGAAAGCA 414  
 QY 416 -GTGAATACTATAAAACATTTTGCATTATCTCGTAAAGATGCAATGGAAGATGACAGCA-- 471  
 Db 415 CATAGATATTATGCTCTGTCCCATCTTCTCGGACACAGACACAGATAGATGGTGCAGGC 474  
 QY 472 -----TATGTAGTACTTAAAAATGACGGCATAACTTTTATGTCAATTTGATGGTTAATACT 525  
 Db 475 AGTGGCTGTGCTTTCTAAATAAATGAAGGACTACTTGTATAAATCATTTATGCTGAACGCA 534  
 QY 526 TGCTATACATACAGCTGAAGGAGTATCTTTGCTACCATATGCTATGCAAGGTATAGGA 585  
 Db 535 TGTATGATGATTAATAGTGAAGGCATACCTTTTCTCTTATATATGTCAGGATTTGGT 594  
 QY 586 GCAGATCTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAGGAAAAATA 645  
 Db 595 ATTGAATTTAGTATCCATGTTTGAAGCTATTAATCTTAAATTTCTTATCAAGGAAATTA 654  
 QY 646 GGTATTAGTTACCTCATCACACAGAAAGTCTCTGCAATTTTATTTGGTGATACTACCATGGC 705

Db 655 GCTTAAGTACCTATAAGCCAGAGCTCTGTGTTATTTGGTGACATTTTCATAAG 714  
 QY 706 GTTATTGGTAATAATTTGAGAAGATACCTGTAAATACCTCGTGTAGTATTAATGATGCT 765  
 Db 715 GTGATAGGAACGAATTTAGAGATATTCCTACTATGATACCTAGTGAATCAGCGCTTGA 774  
 QY 766 CCTCA--AACCACATCTGCTTCACTTCACTTGTGAGTTGGATACCTTTGGCGGAGAAATT 822  
 Db 775 GGAAAGGAATACTACCTCGCAATGTAAACACTGGAGCTGTCTACTTTGGCATAGAACTT 834  
 QY 823 GGAATGAGGTTCACTTC 840  
 Db 835 GGAGGAAGGTTAACTTC 852

## RESULT 18

ID AAC68703 standard; DNA; 861 BP.

XX AAC68703;  
 AC AAC68703;

DT 02-MAR-2001 (first entry)

DE Ehrlichia chaffeensis VSA2 gene partial coding sequence.

XX Ehrlichia chaffeensis; VSA2; variable surface antigen 2; MAPI;  
 KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;  
 KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;  
 KW 3gdorf3; ds.

XX Ehrlichia chaffeensis.

XX WO200605063-A2.

XX 02-NOV-2000.

XX 21-APR-2000; 2000WO-US10886.

XX 22-APR-1999; 99US-0130725.

XX (UYEL) UNIV FLORIDA.

XX Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;  
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;

XX WPI; 2000-679675/66.  
 DR P-PSDB; AAB36186.

XX New polynucleotides useful as DNA vaccines for conferring immunity to  
 PT rickettsial infection e.g. heatwater caused by Cowdria ruminantium,  
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens

XX Claim 4; Page 39-40; 63pp; English.

XX The present sequence shows a high degree of similarity to the major  
 CC antigenic protein 1 (MAPI) gene of Ehrlichia sp. The MAPI gene may be  
 CC used in a vaccines to protect animals or humans against rickettsial  
 CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,  
 CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response  
 CC protective against the rickettsial pathogen. The nucleic acid vaccines  
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.  
 CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1  
 CC and 3gdorf3 may be used in therapeutic and diagnostic applications. The  
 CC polypeptides are useful for detecting antibodies associated with  
 CC infection by a rickettsial pathogen whilst the polynucleotides may be  
 CC used to detect the presence of rickettsial nucleic acids.

XX Sequence 861 BP; 281 A; 140 C; 168 G; 272 T; 0 other;

Query Match 25.3%; Score 212.8; DB 21; Length 861;  
 Best Local Similarity 57.3%; Pred. No. 9.4e-41;

Matches 492; Conservative 0; Mismatches 342; Indels 24; Gaps 5;  
 QY 1 ATGAATTTATAGAAAATCTAGTAAGAAGCGCTTAATCTCATTAATGTCAATCTTACCA 60  
 Db 3 ATGAACCTCGGAAAATTTTATACCAACTGCATTAACATTAATGTCTTCTTACCT 62  
 QY 61 TATCAGTCTTTTGCAGATCTGTAGGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120  
 Db 63 GGAATATACATTTCTGATCCAGTACAG-----GATGACACATTAAGTGGTAATTTCTAC 116  
 QY 121 ATTAGTCAAGATCAATCCAAAGTATATACATTTTGAAGAAATTTCTGTCT--GAAGAA 177  
 Db 117 ATCAGTGGAAAGTATATGCCAAGCGCTTGCATTTTGGAGTTTTTCTGCCAAGGAAGAA 176  
 QY 178 ACTCCTATTAATGGAACAAATTTCTCACTAAAAAGTTTTTCGCACTTAAGAAAGATGGT 237  
 Db 177 AGAAATACACAGTTGGAGTATTTTGGAAATAGAGCAAGATTGGGATAGATGTGTAATATCT 236  
 QY 238 GATATAACAAAAAGACGATTTTACAAGAGTACTCCAGGCATTTGATTTTCAAAAATAAC 297  
 Db 237 AGAACCATTTTAAAGGATATATTCACCGTTCCAAATTTATTCATTTAAGTATGAATAAT 296  
 QY 298 TTAATATCAGGATTTTCAAGAAATTTTGGTATTTCTATGACGACCAAGCAATAGAACTT 357  
 Db 297 CTATTTTCAAGATTTTGCAGGAGCTATTGGCTACTCAATGGATGGCCCAAGAAATAGAGCTT 356  
 QY 358 GAAGCTGCATATCAACAAATTTTAAACAAAAACCCGATCAATGATGATGATGATGATG-- 415  
 Db 357 GAAGTATCTTATGAAGCATTTGATGTTTAAAAAATCAAGGTAAACAATTAAGAACGAAGCA 416  
 QY 416 -GTGAATACTATAAACATTTTGGCATTTCTCGTAAAGATGCAATGGAAGATCAGCAA--- 471  
 Db 417 CATGATATTAATGCTCTGTCCCATCTTCTCGGCACAGACACAGATAGATGGTGCAGGC 476  
 QY 472 -----TATGATGATCTTAAAAATGACGGCATAACTTTTATGTCATTGATGTTAACTACT 525  
 Db 477 AGTGGCTGTCTTTCTAAATAATGAAGGACTACTTGATAAATCATTTATGCTGAACGCA 536  
 QY 526 TGCTATGACATTACAGCTGAAGGAGTATCTTCTGACCATATGATGATGATGATGATGATG 585  
 Db 537 TGTATGATGATTAAGTGAAGGATACCTTTTCTCTTATATATATGTCAGGATTTGGT 596  
 QY 586 GCAGATCTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGAAAAATA 645  
 Db 597 ATTGAATTTAGTATCAATGTTTGAAGCTATAAATCTTAAATTTCTTATCAAGGAAATTA 656  
 QY 646 GGTATTAGTTACCTATACACACAGAAAGTCTCTGCAATTTATTTGGTGGATATACCATGGC 705  
 Db 657 GGCTTAAGTTACCTATAAGCCAGAGCTTCTGTGTTTATTTGGTGGACATTTTTCATAAG 716  
 QY 706 GTTATTGGTATAAATTTGAGAAGATACCTGTAATACTCCTGTAGTATTAATGATGCT 765  
 Db 717 GTGATAGAAACGAATTTTAGAGATATTCTCTACTATGATACCTAGTGAATCAGCGCTTGA 776  
 QY 766 CCTCA---AACACATCTGCTTTCAGTAACCTCTGACCGTTGGATACTTTTGGCGGAGAAAT 822  
 Db 777 GGAAAGGAATACTACCTGCAATAGTAACTGACGCTGACGCTGTTCTACTTTGGCATAGAACT 836  
 QY 823 GGAATGAGGTTCACTTC 840  
 Db 837 GGAGGAGGTTTAACTTC 854

## RESULT 19

AAAX34742

ID AAAX34742 standard; DNA; 756 BP.

XX AC AAAX34742;

XX 05-JUL-1999 (first entry)

XX DE DNA encoding p28 protein.

XX XX

Qy	646	GGTATTAGTTACCCCTATCACACAGAAAGTCTCTGCATTATTTGGTGGATCTATCCATGGC	705
Db	556	GGTTTAAGCTACTCTATAAGCCAGAGCTTCTGTGTTTATTTGGTGGCCTTTTCATAAG	615
Qy	706	GTTATTTGGTAAATAATTGGAGAATAC---CTGTATAAATCTCCTGTAGTATTTAAATGAT	762
Db	616	GTAATAGGGAACGAATTTAGAGATATTCTCTACTATAATACCTACTGGATCAACACTTGCA	675
Qy	763	GCTCCTCAAAACCATCTGCTTTCAGTAACTCTTCAGCTTGGATACTTTTGGCGGAGAAATT	822
Db	676	GGAAAAGGAACATACCTTGCATAATAGTAACTTGGATGTATGCCACTTTTGGAAATAGA	735
Qy	823	GGAAATGAGGTTACCTTC	840
Db	736	GGAGGAAGGTTGTATTC	753
RESULT 20			
AA34748			
ID	AA34748	standard; DNA; 843 BP.	
XX			
AC	AA34748;		
XX			
DT	05-JUL-1999	(first entry)	
XX			
DE	DNA encoding OMP-1F protein.		
XX			
KW	Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;		
KW	detection; dog; ss.		
XX			
OS	Ehrlichia chaffeensis.		
XX			
PN	WO9913720-A1.		
XX			
PD	25-MAR-1999.		
XX			
PF	18-SEP-1998; 98WO-US19600.		
XX			
PR	19-SEP-1997; 97US-0059353.		
XX			
PR	(OHIS ) UNIV OHIO STATE.		
XX			
PA	Ohashi N, Rikihisa Y;		
XX			
PI	WPI; 1999-254290/21.		
XX			
DR	P-FSDB; AAY06948.		
XX			
PT	Novel outer membrane proteins from Ehrlichia chaffeensis and		
XX	Ehrlichia canis		
XX			
PS	Disclosure; Fig 8A; 55pp; English.		
XX			
CC	The invention provides isolated outer membrane proteins (OMP) from		
CC	Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part		
CC	of the OMP family and consist of proteins OMP-1, -1(B to Z) shown		
CC	in AAY06943-958. The E. canis proteins form part of the P30 family and		
CC	consist of proteins shown in AAY06959-970. The proteins and genes are		
CC	used to detect E. chaffeensis in patients and E. canis in dogs.		
XX			
SQ	Sequence 843 BP; 283 A; 128 C; 159 G; 273 T; 0 other;		
Query Match	24.7%;	Score 207.2;	DB 20; Length 843;
Best Local Similarity	56.2%;	Pred. No. 2e-39;	
Matches 477;	Conservative 0;	Mismatches 353;	Indels 18; Gaps 4
Qy	1	ATGAATTATAGAAGAAATTTCTAGTAAGAAGCGCGTTAATCTCATTAAATGTCATCTTACCA	60
Db	1	ATGAATTCAGAAATTTTTTTTATACAACTACATAGTATCGCTAATGCTCTTCTTACT	60
Qy	61	TATCAGTCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGATATAAAGAGGCTTCTTAC	120
Db	61	GGAAATATCAATTTCTGATGCAGTACAGAACGACAAATGTTGCTGTAAT-----TTCTAT	114

QY 121 ATTAGTCAAGTACAATCCAGTATATACACATTTAGAAAATTTCTCTGCTGAA---CAA 177  
 Db 115 ATCAGTGGGAATATGTATCCCAAGTGTTCACATTTTGGCGTATTTCTGCTAAACAGGAA 174  
 QY 178 ACTCCTATTAAATGGAACAAATTTCTCTCAATAAAAGTTTTCGGACTAAAGAAAGATGGT 237  
 Db 175 AGAATACACACACCGGAGTATTTGGATTAAGCAAGATTTGGATGGCGAGCACAATATCT 234  
 QY 238 GATATAACAAAAAGACGATTTTACAAGAGTAGTCCAGGCAATTTGATTTTCAAAATAAC 297  
 Db 235 AAAAATTTCTCCAGAAATACATTTAAACGTTCCAAATTTATTCATTTAAATATGAAAATAAT 294  
 QY 298 TTAATACAGGATTTTCAGGAGTATTTGTTACTTCTGACCGGACCAAGATAGAACTT 357  
 Db 295 CCATTTCTAGGTTTTCAGGAGTGTGTTTAAATTAATTAATTAATTAATTAATTAATTAAT 354  
 QY 358 GAAGCTGATATCAACAATTTTAATCCAAAAACCGATTAACAATGATCTGATATGGT 417  
 Db 355 GAATGTCTATGAACATTTTGATGTGAAAAACCGAGGTATTAATTAAGAACGATGCT 414  
 QY 418 GAATACTATAACATTTTGCATATCTCGT-----AAAGATGCAATGGAAGATCAGCAA 471  
 Db 415 CACAATATATGCTTTAAACCCATAACACAGTGGGGGAAAGCTAAGCAATGCAGGTGATAAG 474  
 QY 472 TATGTAGTACTTAAAAATGACGGCATACTTTTATGTCATTTGATGGTTAATCTGCTAT 531  
 Db 475 TTTGTTTTTCTAAAAAATGAAGGACTACTTGATATATACATTTATTTGAATGCAATGCTAT 534  
 QY 532 GACATTACAGCTGAAGGAGTATCTTTTCGTAACCATATGATGTGCAAGGTATAGGACAGAT 591  
 Db 535 GATGTAATGAAGGAGTATCTTTCTCTTACATATGTGCAAGGTGTTGGTACTGAT 594  
 QY 592 CTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAATAGGATTT 651  
 Db 595 TTAATATCCATGTTTGAAGCTATAAACCCCTAAATTTCTTATCAAGGAAAGTTAGGTTTG 654  
 QY 652 AGTTACCTATCACACGAGTCTCTGCAATTTATTTGGTGGATACATCAATGCGCTTAT 711  
 Db 655 AGTTACTCCATTAAGCCCAAGAGCTCTGTTTTGTTGGTGACATTTTTCATTAAGGTGATA 714  
 QY 712 GGTATAAATTTGAGAAGATACCTGTAATACCTCTGTAGTATTAATTAATGATGCTCCTCAA 771  
 Db 715 GGAATGAATTCAGAGATATTTCTGCTATGATACCCAGTACCTCACTCTCACAGTAT 774  
 QY 772 ACCACATCTGCTTCAGTACTCTTGACGTTGGATATCTTTGGCGGAGAAATGGAATGAGG 831  
 Db 775 CAC---TTTACTATAGTAACACTAAGTGTATGCCACTTTGGAGTGGAACTTTGGAGGAAGG 831  
 QY 832 TTCACCTT 839  
 Db 832 TTTAACTT 839

RESULT 21

AAC68705

ID AAC68705 standard; DNA; 843 BP.

XX AAC68705;

AC AAC68705;

XX AAC68705;

DT 02-MAR-2001 (first entry)

XX Ehrlichia chaffeensis VSA4 gene partial coding sequence.

XX Ehrlichia chaffeensis; VSA4; variable surface antigen 4; MAP1;

KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;

KW Rickettsia; Cowdria; Anaplasma; map2; 1hworf3; 4hworf1; 18hworf1;

KW 3gdorf3; ds.

XX Ehrlichia chaffeensis.

OS Ehrlichia chaffeensis.

XX WO200065063-A2.

PN

XX

PD 02-NOV-2000.

XX 21-APR-2000; 2000WO-US10886.

PF 22-APR-1999; 99US-0130725.

XX (UYFL) UNIV FLORIDA.

XX Barbet AF, Bowie MV, Ganta RR, Burrridge MJ, Mahan SM, McGuire TC;

XX Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Allemen AR;

XX WPI; 2000-679675/66.

XX P-PSDB; AAB36188.

XX New polynucleotides useful as DNA vaccines for conferring immunity to

XX rickettsial infection e.g. heartwater caused by Cowdria ruminantium,

XX comprises major antigenic protein 1 or 2 gene of rickettsial pathogens

XX Claim 4; Page 41; 63pp; English.

XX The present sequence shows a high degree of similarity to the major

XX antigenic protein 1 (MAP1) gene of Ehrlichia sp. The MAP1 gene may be

XX used in a vaccines to protect animals or humans against rickettsial

XX diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,

XX Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response

XX protective against the rickettsial pathogen. The nucleic acid vaccines

XX can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.

XX Cowdria ruminantium genes designated map 2, 1hworf3, 4hworf1, 18hworf1

XX and 3gdorf3 may be used in therapeutic and diagnostic applications. The

XX polypeptides are useful for detecting antibodies associated with

XX infection by a rickettsial pathogen whilst the polynucleotides may be

XX used to detect the presence of rickettsial nucleic acids.

XX Sequence 843 BP; 282 A; 127 C; 159 G; 275 T; 0 other;

SQ

Query Match 24.7%; Score 207.2; DB 21; Length 843;

Best Local Similarity 56.2%; Pred. No. 2e-39;

Matches 477; Conservative 0; Mismatches 353; Indels 18; Gaps 4;

QY 1 ATGAATTTATAGAAAATTTCTAGTAAGAGCGCGTTAATCTCATTAATGTCAATCTTACCA 60

Db 3 ATGAATTTGCAAAAATTTTATTAACAACACTACATTAGTATCGCTAATGTCTTCTTACCT 62

QY 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAGAACTAATGATAACAAGAGGCTTCTAC 120

Db 63 GGAATATCATTTTCTGATGCAGTACAGAACACAAATGTTGGTGGTAAT-----TTCTAT 116

QY 121 ATTAGTCAAAAGTACAATCCAAAGTATATCACACTTTAGAAAATTTCTCTGCTGAA---GAA 177

Db 117 ATCAGTGGGAATATGTACCAAGTGTTCACATTTTGGCGTATTTCTGCTAAACAGGAA 176

QY 178 ACTCCTATTAAATGGAACAAATTTCTCTCACTAAAAAGTTTTCGGACTAAAGAAAGATGGT 237

Db 177 AGAAATACAACAATCGGAGTATTTGGATTAAGCAAGATTTGGGATGGCAGCACAAATATCT 236

QY 238 GATATAACAAAAAGACGATTTTACAAGAGTAGTCCAGGCAATTTTCAAAATAAC 297

Db 237 AAAAATTTCTCCAGAAAATACATTTTAAACGTTTCCAAATTTATTCATTTAAATATGAAAATAAT 296

QY 298 TTAATATCAGGATTTTTCAGGAGTATTTGGTGTATGAGTGGAGTGGGATGAGAGTAT 357

Db 297 CCATTTCTAGTGTTCAGGAGTGTGTTGTTTATTAATGATGTTCCAGAAATAGAGTTA 356

QY 358 GAAGCTGCATATCAACAATTTTAAATCCAAAAACCGGATTAACAATGATGATGATGATGAT 417

Db 357 GAAATGTCTATGAACATTTGATGTGAAAAACCGGAGTAACTAATAAGAACGATGCT 416

QY 418 GAATACTATAACATTTTGTGATTTATCTCGT-----AAAGATGCAATGGAAGATCAGCAA 471

Db 417 CACAATATTTATGCTTTTAAACCCATAACAGTGGGGGAAAGCTTAAGCAATGCGAGGTGATAG 476

QY 472 TATGTAGTACTTAAAAATGACGGCATAACTTTTATGTCATTTGATGGTGGTTAATGCTTAT 531



Db 477 TTTGTTTCTTAAATAAATGAAGGACTACTTGATATATCACTTATGTGAATGCATGCTAT 536  
Qy 532 GACATTACAGCTGAAGAGTATCTTTTCGTACCATATGCTGCAAGGTATAGGAGCAGAT 591  
Db 537 GATGTAATAAGTGAAGGATACCTTTCTCTTACATATGTCAGGTGTTGGTACTGAT 596  
Qy 532 CTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAATAGGTATT 651  
Db 537 TTAATATCCATGTTTGAAGCTATAAACCTTAAATTTCTTATCAAGGAAAATAGGTATT 656  
Qy 652 AGTTACCTATCACACAGAGTCTCTGCATTTATTTGGTGATACCATGCGGTATT 711  
Db 657 AGTTACTCCATAAGCCCAAGGCTTCTGTTTGTGTTGGTGACATTTTCATAAGGTGATA 716  
Qy 712 GGTATAAATTTGAGAAGATACCTGTAATACTCTGTAGTATTAAATGATGCTCTCAA 771  
Db 717 GGGATGAATTCAGAGATATTCCTGCTATGATACCCAGTACCTCACTCACAGGTAA 776  
Qy 772 ACCACATCTGCTTCAAGTAACTCTTGACCTTGATCTTTGGCGGAGAAAATGGAAATGAGG 831  
Db 777 CAC---TTTACTATAGTAACACTAAGTGTATGCCACTTTGGAGTGGAACTTTGGAGGAAG 833  
Qy 832 TTCACCTT 839  
Db 834 TTTAACTT 841

RESULT 22  
AA334747  
ID AA334747 standard; DNA; 837 BP.  
XX AC AA334747;  
XX DT 05-JUL-1999 (first entry)  
XX DE DNA encoding OMP-1E protein.  
XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
XX KW detection; dog; ss.  
XX OS Ehrlichia chaffeensis.  
XX FN WO9913720-A1.  
XX PD 25-MAR-1999.  
XX PF 18-SEP-1998; 98WO-US19600.  
XX PR 19-SEP-1997; 97US-0059353.  
XX PA (OHIS ) UNIV OHIO STATE.  
XX PI Ohashi N, Rikihisa Y;  
XX DR WPI; 1999-254290/21.  
XX DR P-PSDB; AAY06947.  
XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
XX PT Ehrlichia canis  
XX PS Disclosure; Fig 7A; 55pp; English.  
XX CC The invention provides isolated outer membrane proteins (OMP) from  
XX CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
XX CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
XX CC in AAY06943-958. The E. canis proteins form part of the P30 family and  
XX CC consist of proteins shown in AAY06959-970. The proteins and genes are  
XX CC used to detect E. chaffeensis in patients and E. canis in dogs.  
SQ Sequence 837 BP; 280 A; 130 C; 159 G; 268 T; 0 other;  
Query Match 24.4%; Score 205; DB 20; Length 837;

Best Local Similarity 55.3%; Pred. No. 6.5e-39;  
Matches 466; Conservative 0; Mismatches 365; Indels 12; Gaps 3;  
Qy 1 ATGAATTTATAGAAAAATTTCTAGTAAGAGCGGTTAAATCTCATTAATGTCTTAATCTTACCA 60  
Db 1 ATGAATTTGCAAAAAATTTTATTAACAACCTGATAGTATCATTAATGTCTTCTTCTACCT 60  
Qy 61 TATCAGCTTTTTCAGATCTCTGATGTTCAAGAACTAATGATATAACAAGAGGCTTCTTAC 120  
Db 61 GGAATATCATTTTCTGATCCAGTGCA-----AGGTGACAATATTAGTGGTAATTTCTAT 114  
Qy 121 ATTAGTGCAGAGTACAAATCCAGTATATACACATTTTGAAGAAATTTCTCTGCTGAAGAACT 180  
Db 115 GTTAGTGCAGAGTATATGCCAAGTCTTCCGCAATTTTGGCATTTTCTGCTGCAAGAGAA 174  
Qy 181 CCTATTATGGAACAATTTCTCTCACTAAAAAAGTTTTCGAGCTAAAGAAAGATGGTGAT 240  
Db 175 AAAAATCTACTGTGTGATGTTATGGCTTAAACAAGATGGGAAGGATAGCTCATCA 234  
Qy 241 ATAACAAAAAAGACGATTTTACAAGAGTAGCTCCAGGCAATTTTCAAAAATAAATCTTA 300  
Db 235 AGTCACAATGATATCATTTTCAATAACAAGGGTTATTCAATTTAAATATGAAATAACCCA 294  
Qy 301 ATATCAGGATTTTCAGGAAGTATTTGGTACTCTATGACGACCAAGAAATAGAACTTGAA 360  
Db 295 TTTTATAGGTTTTCAGGAGCTATTTGGTTATTCAATGGGTGGTCCAGAGTAGAGTTTGA 354  
Qy 361 GCTGCATATCAACAATTTTAAATCCAAAAACACCGATAACATGATCTGATAATGGTGAA 420  
Db 355 GTGCTCTATGAACATTTTGAGCTTAAATCAGGGTAAATTAATTAATAAATGATGCTCAC 414  
Qy 421 TACTATAAACATTTTGCATTTATCTCGTAAAGATCAATG---GAAGATCAGCAATATGTA 477  
Db 415 AGATACTGTGCTTTTAGGTCAACAGACACAGCGGATACCTTAAACTAGTAAATACGTA 474  
Qy 478 GTACTTAAAAATGACGCATAACTTTTATGTCATGATGGTTAAATCTTGTCTATGACATT 537  
Db 475 CTGTTAAAAAGCGAAGGATTTGCTTGACATATCATTTATGCTAAATGATGCTATGATATA 534  
Qy 538 ACAGCTCAAGGAGTATCTTTGCTACCATATGCTGACGATATAGGAGCAGATCTTATC 597  
Db 535 ATAAACGAGAGTACCTTTGCTCTCTTACATATGTCAGGTGTTGGTACTGATTTAATA 594  
Qy 598 ACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAGGAAAAATAGGTATTAGTTAC 657  
Db 595 TCCATGTTGAAGCTACAAATCTTAAATTTTACCAGGAGATTTAGTCTAAGTTAC 654  
Qy 658 CTTATCACACAGAGTCTCTGCAATTTTATGGTGATACCTACCATGGCGTTATTTGGTAAT 717  
Db 655 TCTATAAACCCAGAAAGCTTCTGTATTTATTTGGTGACATTTTTCATAAGGTGATAGGAAC 714  
Qy 718 AAATTTGAGAAGATACCTGTAATACTCTGTTAGTATTAAATGATGCTCTCAAAACACA 777  
Db 715 GAATTTAGGACATTTCTCTGAAAGCATTTGTTACGTCATCAGTACTCTCCAGATCTA 774  
Qy 778 TCTGCTTCAGTAACTCTTGACGTTGGATCTTTGGCGGAGAAATTTGGAATGAGGTTCACC 837  
Db 775 ---GCAATAGTACACTAAGTGTATGTCATTTTGGATAGACTTTGGAGGAGGTTTAAAC 831  
Qy 838 TTC 840  
Db 832 TTC 834  
RESULT 23  
AAC68704  
ID AAC68704 standard; DNA; 837 BP.  
XX AC AAC68704;  
XX DT 02-MAR-2001 (first entry)  
XX DE Ehrlichia chaffeensis VSA3 gene partial coding sequence.

XX Ehrlichia chaffeensis; USA3; variable surface antigen 3; MAP1;  
 KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;  
 KW Rickettsia; Cowdria; Anaplasma; map2; 1hworf3; 4hworf1; 18hworf1;  
 KW 3gdorf3; ds.  
 XX  
 OS Ehrlichia chaffeensis.  
 XX  
 PN WO200065063-A2.  
 XX  
 PD 02-NOV-2000.  
 XX  
 PF 21-APR-2000; 2000WO-US10886.  
 XX  
 PR 22-APR-1999; 99US-0130725.  
 XX  
 PA (UYFL ) UNIV FLORIDA.  
 XX  
 PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;  
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmore WW, Alleman AR;  
 XX  
 DR P-PSDB; AAB36187.  
 XX  
 DR  
 XX  
 PT New polynucleotides useful as DNA vaccines for conferring immunity to  
 PT rickettsial infection e.g. heatwater caused by Cowdria ruminantium,  
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens  
 PT  
 XX  
 PS Claim 4; Page 40; 63pp; English.  
 XX  
 CC The present sequence shows a high degree of similarity to the major  
 CC antigenic protein 1 (MAP1) gene of Ehrlichia sp. The MAP1 gene may be  
 CC used in a vaccines to protect animals or humans against rickettsial  
 CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,  
 CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response  
 CC protective against the rickettsial pathogen. The nucleic acid vaccines  
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.  
 CC Cowdria ruminantium genes designated map 2, 1hworf3, 4hworf1, 18hworf1  
 CC and 3gdorf3 may be used in therapeutic and diagnostic applications. The  
 CC polypeptides are useful for detecting antibodies associated with  
 CC infection by a rickettsial pathogen whilst the polynucleotides may be  
 CC used to detect the presence of rickettsial nucleic acids.  
 XX  
 SQ Sequence 837 BP; 279 A; 130 C; 159 G; 269 T; 0 other;

Query Match 24.4%; Score 205; DB 21; Length 837;  
 Best Local Similarity 55.3%; Pred. No. 6.5e-39;  
 Matches 466; Conservative 0; Mismatches 365; Indels 12; Gaps 3;

QY 1 ATGAATTATAAGAAATTTCTAGTAGAAGCGGTTAATCTCAATTAATGTCATCTTACCA 60  
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 3 ATGAATTGCAAAAAATTTTATAACAACCTGCATTAGTATCATCAATGTCCTTCTACCT 62  
 QY 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAGAACTAATGATAACAAGAGGCTTCTAC 120  
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 63 GGAATATCATTTTCTGATCCAGTGCA-----AGTGCAATATAGTGGTAATTTCTAT 116  
 QY 121 ATTAGTGCAAGTACAAATCCAAAGTATATCACACTTTAGAAAAATTTCTCTGCTGAAGAACT 180  
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 117 GTTAGTGCAAGTATATGATGCAAGTGTCTGCAATTTTGGCATGTTTTCGCCAAGAGAA 176  
 QY 181 CCTATTATGGACAAATTTCTCTCACTAAAGTTTTCGGACTAAGAAAGATGGTAT 240  
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 177 AAAAAATCCTACTGTGTGTCATTTGATGCTTTAAACACAGATTGGGAAGGATAGTCATCA 236  
 QY 241 ATAACAACAAAAAGACGATTTTACAAGAGTAGTCCAGGCGATTGATTTTCAAAATAACTTA 300  
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 237 AGTCACATATGATATCATTTTCAATACAAAGGTTATTCATTAATATGAAATATACCA 296  
 QY 301 ATATCAGGATTTTCAGGAAGTATTTGGTTACTCTATGGACGGACCAAGATAGAACTTGA 360  
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 297 TTTTATGGGTTTGCAGGAGCTATTGGTTATTCAATGGGTGGTCCAAGAGTAGAGTTGAA 356

QY 361 GCTCATATCAACAATTTAATCCAAAAACACCGATACCAATGATGATGATAATGGTGA 420  
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 357 GTGCTCTTAAGAACATTTGACGTTAAAAATCAGGGTAATACTATAAAAAATGATGCTCAC 416  
 QY 421 TACTATAAACATTTTGCATTTATCTCGTAAAGATGCAATG---GAAGATCAGCAATATGTA 477  
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 417 AGATACCTGTGCTTTAGGTCAACAGACACACAGCGGAATACCTAAAACTAGTAATAACGTA 476  
 QY 478 GTACTTAAAAATAGCGCATAACTTTTATGTCATTGATGGTTAATATCTTGTCTATGACATT 537  
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 477 CTGTTAAAAAGCGAAGGATTCCTTGACATATCAATTTATGCTAAATGCAATGCTATGATATA 536  
 QY 538 ACAGCTGAAGGAGTATCTTTGTCACCATATGTCAGTGCAGGTATAGGAGCAGATCTTATC 597  
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 537 ATNAACGAGAGCATACCTTTGTCCTTACATATGTGCGAGGTGTGGTACTGATTTAATA 596  
 QY 598 ACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAAATAGGTATTAGTTTAC 657  
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 597 TCCATGTTGAGCTACAAATCCTTAAATTTCTTACCAGGGAAGTTAGTCTAAGTTTAC 656  
 QY 658 CCTATCACACAGAAAGTCTCTGCATTTTATGTTGGTGGATCTACCATGGCGTTATTGGTAAAT 717  
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 657 TCTATAAACCCAGAAAGCTTCTGTATTTATTGTTGGTGGACATTTTCTAAGGTGATAGGAAC 716  
 QY 718 AAATTTGAGAGATACCTGTAAATTAACCTCTGTAGTATTAAATGATGCTCTCAAAACCACA 777  
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 717 GAATTTAGGGACATTCCTCTCTGAAAGCATTTGTTAGTCATCAGCTACTCCAGATCTA 776  
 QY 778 TCTGCTTCAGTAACTCTTGACGTTGGATACCTTTGGCGGAGAAAAATTTGGAATCAGGTTCCACC 837  
 DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 777 ---GCAATAGTAACTAAGTGTATGTCATTTTGGAAATAGAACTTTGGAGGAAGGTTTAAAC 833  
 QY 838 TTC 840  
 DB |||||  
 834 TTC 836  
 XX  
 RESULT 24  
 AAV07177  
 ID AAV07177 standard; DNA; 842 BP.  
 XX  
 AC AAV07177;  
 XX  
 DT 14-SEP-1998 (first entry)  
 XX  
 DE Ehrlichia chaffeensis MAP1 gene coding sequence.  
 XX  
 KW MAP1 gene; major antigenic protein 1; rickettsia; DNA vaccine; ss.  
 XX  
 OS Ehrlichia chaffeensis.  
 XX  
 PN WO9816554-A1.  
 XX  
 PD 23-APR-1998.  
 XX  
 PF 17-OCT-1997; 97WO-US19044.  
 XX  
 PR 17-OCT-1996; 96US-0733230.  
 XX  
 PA (UYFL ) UNIV FLORIDA.  
 XX  
 PI Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;  
 PI Nyika A, Rurangirwa FR;  
 XX  
 DR WPI; 1998-251232/22.  
 DR P-PSDB; AAW51089.  
 XX  
 PT Composition containing nucleic acid encoding rickettsial antigen -  
 PT useful for, e.g. stimulating protective immune response in humans or  
 XX animals  
 PS Claim 4; Page 17-18; 39pp; English.





Db 709 GTAATAGGAACGAATTTAGAGATATCTCTACTATAATACCTACTGGATCAACTTGCA 768  
 Qy 763 GCTCCTCAACACCATCTGCTTTCAGTAATCTTTCAGCGTTGGATATCTTTGGCGGAGAAATT 822  
 Db 769 GGAAAGGAACACTACCTCGCAATAGTATAGTATGATGATGATGCCACTTTGGATAGAAATG 828  
 Qy 823 GGAATGAGGTTCA 835  
 Db 829 GGAGGAAGGTTTA 841  
 RESULT 27  
 AAX34745  
 ID AAX34745 standard; DNA; 843 BP.  
 AC AAX34745;  
 DT 05-JUL-1999 (first entry)  
 DX DNA encoding OMP-1C protein.  
 DE Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 KW detection; dog; ss.  
 XX Ehrlichia chaffeensis.  
 OS WO9913720-A1.  
 FN 25-MAR-1999.  
 PD 18-SEP-1998; 98WO-US19600.  
 PF 19-SEP-1997; 97US-0059353.  
 PR (OHIS ) UNIV OHIO STATE.  
 PA Ohashi N, Rikihisa Y;  
 PI WPI; 1999-254290/21.  
 DR P-PSDB; AAY06945.  
 DX Novel outer membrane proteins from Ehrlichia chaffeensis and  
 Ehrlichia canis  
 Disclosure; Fig 5A; 55pp; English.  
 The invention provides isolated outer membrane proteins (OMP) from  
 Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 in AAY06943-958. The E. canis proteins form part of the P30 family and  
 consist of proteins shown in AAY06959-970. The proteins and genes are  
 used to detect E. chaffeensis in patients and E. canis in dogs.  
 Query Match 23.6%; Score 198.4; DB 20; Length 843;  
 Best Local Similarity 54.7%; Pred. No. 2.3e-37;  
 Matches 463; Conservative 0; Mismatches 371; Indels 12; Gaps 3;  
 Qy 1 ATGAATTATAGAAAATCTAGTAGAGAGCGGTTAATCTCAATTAATGCTCAATCTTACCA 60  
 Db 1 ATGAAGCTGCAAAAATTTTATATACCACTGCAATGGCAATGCCAATGCTTTCTTACCT 60  
 Qy 61 TATCAGTCTTTGAGATCCCTGAGGTTCAAGAACTAATGATACAAAGAGGCTTCTAC 120  
 Db 61 GGAATATTACTTTCTGAACCACT-----ACAAGATGACAGTGTGAGTGGCAATTTCTAT 114  
 Qy 121 ATTAGTGCAAAGTACAATCCCAAGTATATCAACATTTAGAAAATTTCTGCTGGAAGAACT 180  
 Db 115 ATTAGTGCAGATACATGCGCAAGTCTTCTCATTTTGGAGTTTCTCTGCGCAAGAGAA 174  
 Qy 181 CCTATTATGAACAAATTTCTCTCACTAAAGAAAGTTTTTCGAGCTAAAGAAAGATGGTGTAT 240

Db 175 AAAAATCCTACTGTGCGGTTGTATGGTTTGAACAAGATGGAAACGGTGTAGTGTCTCA 234  
 Qy 241 ATAAACAAAAAAGACGATTTTTTACAAGAGTAGCTCCAGGCATTGATTTTCAAAAATAACTTA 300  
 Db 235 AGTCATGCTGATGGGACTTTTAATACAAAGTTTATCTTTTAATACGAAAACATATCCA 294  
 Qy 301 ATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGACCGGACCAAGAAATAGAACTTCAA 360  
 Db 295 TTTCTAGGTTTTCAGGAGCTATTGGTTATTTCAATGGTGGTCCAAAGAAATAGAGTTTCAA 354  
 Qy 361 GCTGCATATCAACAATTTAATCCAAAAACACCCGATACACATGATGATGATAATGGTGA- 419  
 Db 355 GTGTCCTATGAACATTTTGACGTTGAAAAATCAAGGTGGTAAATTAACAAAAATGATGCTCAC 414  
 Qy 420 --ATACTATAAACATTTTGCATTATCTCGTAAAGATGCAATGGAAGATACAGCAATATGTA 477  
 Db 415 AGATACTGTGCTTAGATCGTAAAGCAAGCAGCACTAATGCCACAGCTAGTCACTACGTG 474  
 Qy 478 GTACTTAAAAATGACGCATAAATTTTATGTCATTGATGGTTAATATCTGCTATGACATT 537  
 Db 475 CTACTAAAAATGAAGGACTACTTTGATATATCACTTATGTTGAATGCATGCTATGACGTA 534  
 Qy 538 ACAGCTGAGGAGTATCTTTGCTACCATATGCTGTCAGGTATAGGAGGAGATCTTATC 597  
 Db 535 GTAAGTGAAGGAATACCTTTCTCTCTTACATATGTGCAGGTGTGGTACCGATTTAATA 594  
 Qy 598 ACTATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGAAAAAATAGGTATTAGTTAC 657  
 Db 595 TCCATGTTTGAAGCTATAAACCTTAATTTCTTATCAAGGAAAGTATAGGTTTGAAGTTAC 654  
 Qy 658 CCTATCACACAGAAAGTCTCTGCATTTTATGGTGGATACCTACCATGGCGTTATTGGTAAAT 717  
 Db 655 TCTATAAACCCAGAAAGCTTCTGCTTTGTTGGTGACATTTTTCATAAAGTTGCAAGTAAAT 714  
 Qy 718 AAATTTGAGAGAT---ACCTGTAATACTCCTGTAGTATTAATGATGCTCTCTCAAAACC 774  
 Db 715 GAATTCAGGGACATTTCTACTCTTAAAGCGTTTGTACACCATCATCTGCGAGCTACTCCA 774  
 Qy 775 ACATCTGCTTCAGTAACTCTTGACGTTGGATACCTTGGCGAGAAATTTGGAATGAGGTTTC 834  
 Db 775 GACTTAGCAACAGTAACACTGAGTGTGTGTCACATTTGGAGTAGAAGAACTTGGAGGAAGATTT 834  
 Qy 835 ACCTTC 840  
 Db 835 AACTTC 840  
 RESULT 28  
 ABK68877  
 ID ABK68877 standard; DNA; 828 BP.  
 AC ABK68877;  
 XX 02-JUL-2002 (first entry)  
 DT DNA encoding Ehrlichia canis p28-3.  
 DE Ehrlichia canis infection; vaccine; serodiagnostic; gene; p28; ss;  
 KW antibacterial.  
 XX Ehrlichia canis.  
 OS WO200222782-A2.  
 FN 21-MAR-2002.  
 PD 12-SEP-2001; 2001WO-US28759.  
 PF 12-SEP-2000; 2000US-0660587.  
 PR (RERE-) RES DEV FOUND.  
 XX

PI Walker DH, Yu X, McBride JW;  
XX WPI; 2002-351882/38.  
DR P-PSDB; AAU96117.  
XX New recombinant homologous 28 kilodalton immunodominant protein from  
PT Ehrlichia canis, useful for treating Ehrlichia canis infections -  
XX Claim 5; Figure 15; 106pp; English.  
XX The invention relates to a recombinant homologous 28 kDa immunodominant  
CC protein, p28 (I) of Ehrlichia canis. (I), a 28-kDa antigen preferably  
CC dispersed in a pharmaceutically acceptable carrier, is useful for  
CC inhibiting E. canis infection in a subject. (I) is useful in the  
CC development of vaccines and serodiagnostics that are particularly  
CC effective for disease prevention and serodiagnosis. ABK68852-ABK68878  
CC represent the 28-kDa antigen coding sequences and PCR primers of the  
CC invention.  
XX Sequence 828 BP; 300 A; 126 C; 136 G; 266 T; 0 other;  
SQ  
Query Match 23.2%; Score 195; DB 24; Length 828;  
Best Local Similarity 53.9%; Pred. No. 1.5e-36;  
Matches 452; Conservative 0; Mismatches 375; Indels 12; Gaps 2;  
QY 1 ATGAATTATAGAAAATCTAGTAGAAGCGGTTAATCTCAATTAATGTCATCTTACCA 60  
DB 1 ATGAAGCTGTAATAAATCTTATATAACAATACATACATGTTGTTATCTTTTACCT 60  
QY 61 TATCAGTCTTTGGCAGATCCTGTAGGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120  
DB 61 GGCATATCTTCCAAACCAATACATGAA-----AACATCTACAGAACTTTTAC 114  
QY 121 ATTAGTGCAGAGTACAACTCAAGTATATACACTTTAGAAAATCTCTGCTGAAGAACT 180  
DB 115 ATTTTGGAAAATATGTACCAAGTATTTTCACTTTTGGGAATCTTTTACGCTAAAGAAGAA 174  
QY 181 CCTATTATGGAACAAATCTCTCAATAAAGATTTTCGGACTAAAGAGATGGTAT 240  
DB 175 AAAACACAAACAACTGGAATTTTGGATTTAAAGAACTCATGCACTGGTGTATCATCTT 234  
QY 241 ATACAAAAGAGAGATTTTACAAGAGTAGCTCCAGGCAATGATTTTCAAAATAACTTA 300  
DB 235 GATAAGAACATGAGCTTTTAAATATCCAAATTTATTTAAATATGAATATATCA 294  
QY 301 ATATCAGGATTTTCAGGAAGTATTTGTTACTCTATGACGGACCAAGAAATAGAACTTGA 360  
DB 295 TTTTATGAGATTTTCAGGGGTAATTTGGCTATTCAATAGGTAGTCCCAAGAAATAGAA 354  
QY 361 GCTGCATATCAAAATTTAATCCAAACACCGATACAAATGATGATGATGATGATGATG 420  
DB 355 GTATCATACGAGACATTCGATGTACAAATCCAGGAGATAGTTTAAACAATGATGCAT 414  
QY 421 TACTATAAATTTTGCATTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGATGTA 480  
DB 415 AGATATGCTTTATCCATGATTCAGTAAACAAATGAAAGTGGTAAATTCGTTTT 474  
QY 481 CTTAAAAATGACGGCAATACCTTTTATGCTATGATGTTTAAATCTTGTATGATGATGAT 540  
DB 475 CTTAAAAATGAGGATTAAGTGACATATCACTCATGTTAAATGATGTTATGATGATGAT 534  
QY 541 GCTGAGGAGTATCTTTCGTACCATATGATGCTGAGGTATAGGAGCAGATCTTCACT 600  
DB 535 AACAAAAGAAATGCCCTTTTTCACCTTACATATGTCAGGCAATTTGTTACTGACTTAATTC 594  
QY 601 ATTTTAAAGACCTCAATCTAAAATTTGCTTACCAAGGAAAAATAGGTATTAGTTACCT 660  
DB 595 ATGTTTGAGCTTAAACCATTAAGCTGCTTATCAAGGAAATTAGTTTAAATATCA 654  
QY 661 ATCAACCAAGAGTCTCTGCAATTTATTTGTTGATATCTACCATGCGGTTATTTGGTAATA 720  
DB 655 ATAAGCCCAAGAGCTAAACATTTCTATGGGTGTGCACTTTTCAAAAGATCAACAAACAG 714

QY 721 TTTGAGAAGATACCTGTAAATAAATCCTGTAGTATTAAATGATGCTCCTCAACACACATCT 780  
DB 715 TTTAGAGTTCTGTTCTATTAACTGCTGGAGACT-----CGTCCAGATAATCTATTT 768  
QY 781 GCTTCAGTAACCTTTGACGTTGGATACATTTTGGCGGAGAAATTTGAATGAGGTTACCTT 839  
DB 769 GCAATAGTAAAGTTGAGTATATGTCATTTTGGGTAGAAATTTGGGTACAGGTCAGTTT 827  
RESULT 29  
AAV07176  
ID AAV07176 standard; DNA; 864 BP.  
AC AAV07176;  
XX 14-SEP-1998 (first entry)  
DT  
XX Cowdria ruminantium MAP1 gene coding sequence.  
DE  
XX MAP1 gene; major antigenic protein 1; rickettsia; heartworm;  
KW DNA vaccine; ss.  
OS Cowdria ruminantium.  
XX WO9816554-A1.  
PN  
XX 23-APR-1998.  
PD  
XX 17-OCT-1997; 97WO-US19044.  
PF  
XX 17-OCT-1996; 96US-0733230.  
PR  
XX (UYFL) UNIV FLORIDA.  
PA  
XX Barbet AF, Burrigle MJ, Ganta RR, Mahan SM, McGuire TC;  
PI Nyika A, Rurangirwa FR;  
XX WPI; 1998-251232/22.  
DR P-PSDB; AAW51088.  
XX Composition containing nucleic acid encoding rickettsial antigen -  
PT useful for, e.g. stimulating protective immune response in humans or  
PT animals  
XX Claim 4; Page 14-15; 39pp; English.  
PS  
XX This DNA molecule comprises the coding region of the major antigen  
CC protein 1 gene (MAP1) of Cowdria ruminantium, the causative agent  
CC of heartwater in domestic ruminants. It codes for a 287-amino acid  
CC MAP1 protein (see AAW51088). A claimed composition comprises a  
CC nucleic acid (see AAV07176-82) encoding a polypeptide (see AAW51088-99)  
CC that elicits a protective immune response against a rickettsial  
CC pathogen. The nucleic acid is used, in human or veterinary  
CC medicine, in vaccines to protect against Rickettsia, Ehrlichia,  
CC Anaplasma and Cowdria species. The nucleic acids are also useful  
CC as probes to identify related sequences, e.g. for identification of  
CC organisms and for diagnosing infection. Use of nucleic acid  
CC vaccines avoids the problem of protein purification associated with  
CC protein-based vaccines. The nucleic acid does not replicate in the  
CC host but remains episomal and capable of expressing polypeptide for  
CC at least 19 mth.  
XX  
SQ Sequence 864 BP; 307 A; 139 C; 149 G; 269 T; 0 other;  
Query Match 23.2%; Score 195; DB 19; Length 864;  
Best Local Similarity 55.4%; Pred. No. 1.5e-36;  
Matches 475; Conservative 0; Mismatches 355; Indels 27; Gaps 4;  
QY 1 ATGAATTATAGAAAATCTAGTAGAAGCGGTTAATCTCAATTAATGTCATCTTACCA 60  
DB 1 ATGAATTCGAAGAAAATTTTATCAAGATGATCAATATCATGTTGTTTACCT 60  
QY 61 TATCAGTCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120

```
Db 61 GGTGTCCTTTCTGATGTAATACAGGAACACAGCAACCCAGC---AGGCAGTGTCTTAC 117
Qy 121 ATTAGTCAAAAGTACAATCAAGTATATACACATTTTAGAAAATCTCTGCTGAAGAACT 180
Db 118 ATTAGCGCAATATACATGCGCACTGATCATCATTTTGGTAATATGTCATCAAGAAGAT 177
Qy 181 CCTATTAAATGAACAAATTTCTCTCACTAAAAA-----GTTTTCGGACTAAAG 228
Db 178 TCAAAAAATACTCAACCGTATTTGGTCTAAAAAAGATTTGGATGGGTTTAAACACCA 237
Qy 229 AAAGATGCTGATATAACAAAAAGACGATTTTACAGAGTAGCTCCAGGCAATGATTTT 288
Db 238 TCAGATTTCTAGCAATATAATTTCTACAATTTTACTGAAAAAGACTATTCTTTTCAGATAT 297
Qy 289 CAAAATACCTTAATATCAGGATTTTTCAGGAAGTATTTGTTACTCTATGCGCGACCAAGA 348
Db 298 GAAAAAATCGGTTTTTAGGTTTCGCTGGACCAATTTGGTACTCAATGAATGGACCAAGA 357
Qy 349 ATAGAACTTGAAGCTGCATATCAACAATTTTAAATCCAAAAACACCGATAACAATGATCT 408
Db 358 ATAGAGTTTCGAAGTATCTCTATGAAACTTTTGATGTAAAAAACCTAGGTGGCACTATAAA 417
Qy 409 GATATGCTGATATCTATAACATTT-----TGCATTTATCTCGTAAAGATGCAATG 459
Db 418 AACAAACGCACACATGCTACTGTGCTTTAGATACAGCGCACAAAATAGCACTAATGGCGCA 477
Qy 460 GAAGATCAGCAATATGTAGTACTTTAAATGACGCGCATACCTTTTATGTCATTCATGTT 519
Db 478 GGATTAACATCATCTGTTATGGTAAACGAAAAATTTACAAATATATCATTAATGTTA 537
Qy 520 AATCTGCTATGATATPACAGCTGAAGAGTATCTTTTCGTAACCATATGCAATGTCAGGT 579
Db 538 AATCGGTATGATATCATGCTTGATGGAATACCAAGTTCTCCATATGTTATGTCAGGT 597
Qy 580 ATAGAGCAGATCTATCACTATTTTAAAGACCTCAATCTAAAATTTGCTTTACCAAGA 639
Db 598 ATTGGCACTGACTAGTGTCTAGTAAATTAATGCTACAAATCTCTAAATTTATCTTATCAAGGA 657
Qy 640 AAAATAGTATTTAGTTACCTTATCACACGAAGTCTCTGCAATTTATTTGGTGGATACCTAC 699
Db 658 AAGCTAGCAATAGTTACTCAATCAATCTGAGCTCTATCTTTATCGGTGGACATTC 717
Qy 700 CATCGGCTTATGTTAAATTTGAGAAGATACCT---GTAATAACTCTCTGTAGTATTA 756
Db 718 CATAGAGTTATAGGTAATGAATTTAAAGATATTGCTACCTTAAAAAATATTACTTCAAAA 777
Qy 757 AATGATGCTCTCAACACCATCTGCTTCAGTAACTCTTGACGTTGGATATCTTTGGCGGA 816
Db 778 ACAGGAATATCTAATCTCGGCTTTTGCATCAGCAACACTTGATGTTTGTCACTTTGGTATA 837
Qy 817 GAAATTTGGAATGAGTT 833
Db 838 GAAATTTGGAAGAGTT 854
```

## RESULT 30

AAC68699  
ID AAC68699 standard; DNA; 864 BP.

XX AC AAC68699;

XX DT 02-MAR-2001 (first entry)

XX DE Cowdria ruminatum MAP1 gene.

XX KW Cowdria ruminatum; MAP1; major antigenic protein 1; antirickettsial;  
vaccine; gene therapy; Rickettsia; Ehrlichia; Anaplasma; map2; 1hworf3;  
4hworf1; 18hworf1; 3gdorf3; ds.

XX OS Cowdria ruminatum.

XX FN W0200065063-A2.

```
XX PD 02-NOV-2000.
XX PF 21-APR-2000; 2000WO-US10886.
XX PR 22-APR-1999; 99US-0130725.
XX PA (UYFL) UNIV FLORIDA.
XX PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC,
PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Allenan AR;
XX WPI; 2000-679675/66.
XX DR P-ESDB; AAB36182.
XX PT New polynucleotides useful as DNA vaccines for conferring immunity to
PT rickettsial infection e.g. heartwater caused by Cowdria ruminatum,
PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
XX
XX Claim 4; Fig 1; 63pp; English.
XX
CC The present sequence is given in a specification relating to nucleic
CC acid vaccines containing genes to protect animals or humans against
CC rickettsial diseases caused by a organisms of Rickettsia sp., Ehrlichia
CC sp., Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
CC protective against the rickettsial pathogen. The vaccine comprises the
CC major antigenic protein 1 (MAP1) gene or the major antigenic protein 2
CC (MAP2) gene of rickettsial pathogens. The nucleic acid vaccines can be
CC driven by the human cytomegalovirus (HCMV) enhancer-promoter. Cowdria
CC ruminatum genes designated map 2, 1hworf3, 4hworf1, 18hworf1 and
CC 3gdorf3 may be used in therapeutic and diagnostic applications. The
CC polypeptides are useful for detecting antibodies associated with
CC infection by a rickettsial pathogen whilst the polynucleotides may be
CC used to detect the presence of rickettsial nucleic acids.
XX
XX Sequence 864 BP; 307 A; 139 C; 149 G; 269 T; 0 other;
```

Query Match 23.2%; Score 195; DB 21; Length 864;

Best Local Similarity 55.4%; Pred. No. 1.5e-36;

Matches 475; Conservative 0; Mismatches 355; Indels 27; Gaps 4;

```
Qy 1 ATGAATTTAAGAAAAATTTCTAGTAAGAGCGGCTTAATCTCAATTAATGCTCAATCTTACCA 60
Db 1 ATGAATTTGCAAGAAAAATTTTATCAAGTACACTAATATCATTTAGTGTCTATTTTACCT 60
Qy 61 TATCAGTCTTTTCAGATCTCTGAGGTTCAAGACTAATGATAACAAGAGGCTTCTAC 120
Db 61 GGTGTGCTCTTTCTGATGTAATACAGNAAGACAGCAACCCAGC---AGGCAGTGTCTTAC 117
Qy 121 ATTAGTCAAAAGTACAATCCAAGTATATCACACTTTTAGAAAATTTCTCTGCTGAAGAACT 180
Db 118 ATTAGCGCAAAATATACATGCCAATGCAATCTGTCATCATTTTGGTAAATGTCATCAAGAAGAT 177
Qy 181 CCTATTAAATGAACAAATTTCTCTCACTAAAAA-----GTTTTCGGACTAAAG 228
Db 178 TCAAAAAATACTCAACCGTATTTGGTCTAAAAAAGATTTGGATGGGTTTAAACACCA 237
Qy 229 AAAGATGCTGATATAACAAAAAAGACGATTTTACAGAGTAGCTCCAGGCAATGATTTT 288
Db 238 TCAGATTTCTAGCAATATAATTTCTACAATTTTACTGAAAAAGACTATTCTTTTCAGATAT 297
Qy 289 CAAAATACCTTAATATCAGGATTTTTCAGGAAGTATTTGTTACTCTATGCGCGACCAAGA 348
Db 298 GAAAAAATCGGTTTTTAGGTTTCGCTGGACCAATTTGGTACTCAATGAATGGACCAAGA 357
Qy 349 ATAGAACTTGAAGCTGCATATCAACAATTTTAAATCCAAAAACACCGATAACAATGATCT 408
Db 358 ATAGAGTTTCGAAGTATCTCTATGAAACTTTTGATGTAAAAAACCTAGGTGGCACTATAAA 417
Qy 409 GATATGCTGATATCTATAACATTT-----TGCATTTATCTCGTAAAGATGCAATG 459
Db 418 AACAAACGCACACATGCTACTGTGCTTTAGATACAGCAGCAAAAATAGCACTAATGGCGCA 477
```

QY 460 GAAGATCAGCAATATGTAGTACTTAAATGACGGCATAACTTTTATGTCATTGATGTT 519  
 Db 478 GGATTTAACTACATCTGTTATGTTGTAATAAAGCGAAATTTTAAACAATATATCATTTATGTTA 537  
 QY 520 AATCTCTGTATGACATACAGCTCAAGGAGTATCTTTCGTACCATATGTCATGTCAGGT 579  
 Db 538 AATGCGTGTATGATATCATGCTTGATGGAATACAGTTCTCCATATGTTATGTCAGGT 597  
 QY 580 ATAGGAGAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGA 639  
 Db 598 ATTGSCACTGACTTAGTGTCAATTAATGCTACAAATCTTAAATTTATTTATCAAGGA 657  
 QY 640 AAAATAGTATTAGTTACCTATACACAGAGTCTCTGCAATTTATTTGGTGGATAC 699  
 Db 658 AAGCTAGGCAATAGTTACTCAATCAATTTCTGAAGCTTCTATCTTTATCGGTGGACATTC 717  
 QY 700 CATGCGTTATTGGTAAATAATTTTCAGAGATACCT---GTAATAACTCTGTAGTATTA 756  
 Db 718 CATAGATTATAGGTAAATTTTAAAGATATGCTTAAATTTTATTTACTTCAAAA 777  
 QY 757 AATGATGCTCTCAAAACACACTCTGCTTCAGTAATCTTTGAGCTTTGGATACCTTTGGCGGA 816  
 Db 778 ACAGGAATATCTAATCTGCTGCTTTCATCAGCAACACTTGATGTTTGTCACTTTTGGTATA 837  
 QY 817 GAAATGGAAATGAGGTT 833  
 Db 838 GAAATGGAGGAAGGTT 854

## RESULT 31

ID AAS07575 standard; DNA; 864 BP.

AC AAS07575;

DT 23-OCT-2001 (first entry)

DE DNA encoding major antigenic protein 1 (MAP1).

XX Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;  
 KW infection; heartwater; diagnostic; ds.

XX Cowdria ruminantium.

XX Key Location/Qualifiers

FT 1..864

FT /\*tag= a

FT /product= "Major antigenic protein 1 (MAP1)"

XX US6251872-B1.

PN 26-JUN-2001.

XX 17-OCT-1997; 97US-0953326.

XX 17-OCT-1996; 96US-0733230.

XX (UYFL ) UNIV FLORIDA.

XX Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;

PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;

XX WPI; 2001-424487/45.

DR P-PSDB; AAU04192.

XX New MAP2 genes and polypeptides useful as vaccines for conferring  
 PT immunity to human and animal rickettsial diseases, e.g. heartwater, or  
 PT as molecular markers in nucleic acid analysis procedures -

XX Example 1; Fig 1; 30pp; English.

XX The sequence represents the coding sequence of major antigenic protein 1

CC (MAP1) isolated from Cowdria ruminantium. The MAP polynucleotides and  
 CC polypeptides are useful as vaccines for conferring immunity to rickettsia  
 CC infection, including Cowdria ruminantium causing heartwater. The MAP  
 CC polynucleotides may be used as molecular markers in nucleic acid  
 CC analysis procedures, and to produce the MAP polypeptides, which may  
 CC be used to raise antibodies that are reactive with the polypeptides.  
 CC The nucleic acids may further be used as probes to identify  
 CC complementary sequences within other nucleic acid molecules or genomes,  
 CC where such probes can be applied to identify or distinguish infectious  
 CC strains of organisms in diagnostic procedures or in rickettsial  
 CC research where identification of particular organisms or strains is  
 CC needed.

SQ Sequence 864 BP; 307 A; 139 C; 149 G; 269 T; 0 other;

Query Match 23.2%; Score 195; DB 22; Length 864;

Best Local Similarity 55.4%; Pred. No. 1.5e-36;

Matches 475; Conservative 0; Mismatches 355; Indels 27; Gaps 4;

QY 1 ATGAATTATAGAAATTTCTAGTAAGAGCGGTTAATCTCAATTAATGTCATCTTACCA 60  
 Db 1 ATGAATTGCAAGAAATTTTATCACAAGTACATAATATCATTTAGTGTCAATTTTACCT 60  
 QY 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAGAGAACTAATGATAACAAAGAAAGGCTTCTAC 120  
 Db 61 GGTGTGTCCTTTCTGATGTAATACAGAGACAGCAACCCAGC---AGGCAGTGTTTAC 117  
 QY 121 ATTAGTGCAAGTACAATCCAAGTATATCACATTTTAGAAAAATTTCTGTGGAAGAAACT 180  
 Db 118 ATTAGCGCAAAATACATGCCAACTGCATCACATTTTGTGTAATGTCAATCAAGAAAGAT 177  
 QY 181 CCTATTATGGAACAATTTCTCTCACTAAAAA-----GTTTTCGGACTAAAG 228  
 Db 178 TCAAAAAATACTCAAAACGGTATTTGGTCTAAAAAAGATTTGGGATGGCGTTAAACACCA 237  
 QY 229 AAGATGGTGATATAACAAAAAAGACGATTTTACAGAGTAGTCCAGGCAATGATTTT 288  
 Db 238 TCAGATTTAGCAATACTAATTTCTACAATTTTACTGNAAGAGACTATTTCTCAGATAT 297  
 QY 289 CAAAATACTTAATATCAGGATTTTTCAGGAAGTATTTGGTTACTCTATGGACGGACCAAGA 348  
 Db 298 GAAAAAATCCGTTTTTGGTTCGCTGGAGCAATTTGGGTACTCAATGAATGGACCAAGA 357  
 QY 349 ATAGAACTTGAAGTGCATATCAACATTTTATCCAAAAACACCGATAACAAATGATACT 408  
 Db 358 ATAGAGTTTCAAGTATCTCTATGAAACTTTTATGATAAAAAACCTTAGGTGGCAACTATAA 417  
 QY 409 GATAATGGTGAATACTATAAACATTT-----TGCAATTTATCTCGTAAAGATGCAATG 459  
 Db 418 AACACGACACATGTACTGTGCTTTAGATACAGCAGCACAATAAGCAGCTAATGGCGCA 477  
 QY 460 GAAGATCAGCAATATGTAGTACTTAAATAATGACGGCATAACTTTTATGTCATTTGATGTT 519  
 Db 478 GGATTAACATCATCTGTTATGTTAAACCGAAAAATTTTAACAAAATATATCATTTAATGTTA 537  
 QY 520 AATACTTGTATGACATTTACAGCTGAAGAGTATCTTTCGTACCATATGTCATGTCAGGT 579  
 Db 538 AATCGTGTATGATATCATCTTGTATGGAATACAGTTTCTCCATATGTTATGTCAGGT 597  
 QY 580 ATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGA 639  
 Db 598 ATTGGCACTGACTTAGTGTAGTAAATTAATGCTACAAATCTTAAATTTATCTTATCAAGGA 657  
 QY 640 AAAATAGGTATAGTTACCTATCACACAGAGTCTCTGCAATTTTATTTGGTGGATAC 699  
 Db 658 AAGCTAGGCATAAGTTACTCAATCAATTTCTGAAGCTTCTATCTTTTATCGGTGGACATTC 717  
 QY 700 CATGCGTTATGTTAAATAATTTGAGAGATACCT---GTAATAACTCTCTGTAGTATTA 756  
 Db 718 CATAGATTATAGGTAAATTTTAAAGATATTTGCTACCTTAAATAATTTTACTTCAAAA 777  
 QY 757 AATGATGCTCTCAAAACACATCTGCTTCAGTAATCTTTGAGCTTTGGGATACCTTTGGCGGA 816







Db 115 ATATCAGGAAGTACATGCCAAGTGTCTCTCATTTTGGAAATTTTTCAGCTGAAGAGAG 174  
 Qy 181 CCTATTAAATGAAC---AAATTCCTCTCACTAAAGAAAGTTTTCGGAATAAGAAAGATGGT 237  
 Db 175 AAAAAAAGACACACTGTAGTATATGGCTTAAAGAGAAATCTGGCAGGAGATGCAATATCT 234  
 Qy 238 GATATACAAAAAAGACGATTTTACAGAGTAGTCTCCAGGANTGATTTTCAAAATTAAC 297  
 Db 235 AGTCAAGTCCAGATGATATTTTACCATTTCGAATTTACTCATTTCAAGTATGCAAGCAAC 294  
 Qy 298 TTAATATCAGGATTTTTCAGGAAGTATTTGGTTTACTCTATGGACGACCAAGAAATGAACCTT 357  
 Db 295 AAGTTTTTAGGGTTTCAGTAGTATTTGGTTTACTCTGATAGGAGTCCCAAGAAATAGAATT 354  
 Qy 358 GAAGCTGCATATCAACAATTTTAATCCAAAAACCCGATAACAATGATPACTGA----- 410  
 Db 355 GAGATGCTTATGAAGCATTTTGTGTAAGAAATCCAGGTGATAATTACAAAAACGGTGCT 414  
 Qy 411 -----TAATGGTGAATACATATAAACAATTTTGCATTTCTGTAAGATGCAATGGAAGAT 465  
 Db 415 TACAGGTATTTGGCTTTTATCTCATCAAGATGATCGGATGATGACATGACTAGTGCAACT 474  
 Qy 466 CAGCAATATGATGACTTTAAAAATGACGGCATAACTTTTATGTCATTTGATGTTAAATACT 525  
 Db 475 GACAAATTTGATATTTAATTAATGAAGATTAATTAACATATCATTTATGACAAACATA 534  
 Qy 526 TGCTATGATATCAGCTGAAGAGTAGTCTTTCGTACCATATGATGCTGAGGATATAGA 585  
 Db 535 TGTATGAAACAGCAAGCAAAAAATATACCTCTCTCTCTTACATATGTCGAGGTATTCGT 594  
 Qy 586 GCAGATCTTATCACTATTTTAAAGACCTCAATCTAAAATTTGCTTACCAAGGAAAAATA 645  
 Db 595 ACTGATTTAATTCACATGTTTGAAGTACACATCTTAAATTTCTTATCAAGGAAAGCTA 654  
 Qy 646 GGTATTAGTTACCTCATCACACCAAGAGTCTCTGCAATTTATGTTGGATATCACTACG 705  
 Db 655 GGGTGGCTACTTCGTAAGTGCAGAGTCTTCGGTTCTTTTGGTATATATTTTCATAAA 714  
 Qy 706 GTTATTGGTATAAATTTGAGAAGATACCTGTAAATACCTCCTGTAGTATTAATATGATGCT 765  
 Db 715 ATTTAATAATAAGTTTAAAAATGTTCCAGCCATGGTACCTATTAATCAAGAGATA 774  
 Qy 766 CTTCAACCAACATCTGCTTCAAGTAACTCTTCAAGTTGGATATCTTTGGCGGAGAAATGGA 825  
 Db 775 GTAGACCAACAGTTTGCACAGATTAACATTAATATGATGCTACTTTTGGATTAGAATTGGA 834  
 Qy 826 ATGAGGTTTCACTTC 840  
 Db 835 TGTAGTTCAACTTC 849

## RESULT 35

ABK68853

ID ABK68853 standard; DNA; 849 BP.

XX AC

XX ABK68853;

XX DT

XX 02-JUL-2002 (first entry)

XX DE

XX DNA encoding Ehrlichia canis p28-5.

XX XX

XX Ehrlichia canis infection; vaccine; serodiagnostic; gene; p28; ss;

XX KW

XX antibacterial.

XX OS

XX Ehrlichia canis.

XX XX

XX WO200222782-A2.

XX FN

XX 21-MAR-2002.

XX PD

XX 12-SEP-2001; 2001WO-US28759.

XX PF

XX 12-SEP-2000; 2000US-0660587.

XX PR

(RERE-) RES DEV FOUND.

Walker DH, Yu X, McBride JW;

WPI; 2002-351882/38.

P-PSDB; AAU96101.

New recombinant homologous 28 kilodalton immunodominant protein from Ehrlichia canis, useful for treating Ehrlichia canis infections -

Claim 5; Figure 7; 106pp; English.

The invention relates to a recombinant homologous 28 kDa immunodominant protein, p28 (I) of Ehrlichia canis. (I), a 28-kDa antigen preferably dispersed in a pharmaceutical acceptable carrier, is useful for inhibiting E. canis infection in a subject. (I) is useful in the development of vaccines and serodiagnostics that are particularly effective for disease prevention and serodiagnosis. ABK68852-ABK68878 represent the 28-kDa antigen coding sequences and PCR primers of the invention.

SQ Sequence 849 BP; 293 A; 131 C; 152 G; 273 T; 0 other;

Query Match 21.6%; Score 181.4; DB 24; Length 849;  
 Best Local Similarity 53.6%; Pred. No. 2.4e-33;

Matches 458; Conservative 0; Mismatches 376; Indels 21; Gaps 3;

Qy 1 ATGAATTATAGAAATCTAGTAAGAGCGGTTAATCTCAATTAATGCTCAATCTTACCA 60  
 Db 1 ATGAATTGTAAAAAGTTTTCACAAATAGTGCATGATATCATATCATATCTTCTACT 60  
 Qy 61 TATCAGCTCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120  
 Db 61 AATGCTCATACTCTACCCAGTATATGGTAACAGTATGATGTGTAAT-----TTTAC 114  
 Qy 121 ATTAGTCAAAAGTACAATCCAAAGTATATCACATTTTGAATAATCTCTGCTGAAGAACT 180  
 Db 115 ATATCAGAAAGTACATGCCAAGTGTCTCTCATTTTGGAAATTTTTCAGCTGAAGAGAG 174  
 Qy 181 CCTATTATGGAAC---AAATTCCTCACTAAAAGTTTTCGGNCTAAAGAAAGATGGT 237  
 Db 175 AAAAAAAGCACTGTAGTATATGGCTTAAAGAAACTGGGCGAGGAGATGCAATATCT 234  
 Qy 238 GATATAACAAAAAGACGATTTTACAAAGTAGTCTCCAGGCATTTGATTTTCAAAATAAC 297  
 Db 235 AGTCAAGTCCAGATGATATTTTACCAATTCGAATTTACTCATTTCAAGTATGCAAGCAAC 294  
 Qy 298 TTAATATCAGGATTTTTCAGGAAGTATTTGGTTACTCTATGGACGGACCAAGAAATAGAACTT 357  
 Db 295 AAGTTTTTAGGGTTTGCAGTAGCTATTTGGTTTACTCGATAGGCAGTCCCAAGAAATAGAAGTT 354  
 Qy 358 GAAGCTGCATATCAACAATTTAATCCAAAAACCCGATAACAATGATGACTGA----- 410  
 Db 355 GAGATGCTTATGAAGCATTTGATGTGAAAAATCCAGGTGATATTAATTAACAAAAACGGTGCT 414  
 Qy 411 -----TAATGGTGAATATAATAACATTTTGCATTTATCTCGTAAAGATGCAATGGAAGAT 465  
 Db 415 TACAGGTATTTGGCTTTTATCTCATCAAGATGATCGGATGATGATGATGCACTAGTGCAACT 474  
 Qy 466 CAGCAATATGATGATCTTAAAAATGACGGCATAAATTTTATGTCATGATGCTTAAATACT 525  
 Db 475 GACAAATTTGTATATTTAATTAATGAAGGATTACTTAAACATATCATTTATGACAAACATA 534  
 Qy 526 TGCTATGACATTAACAGCTGAAGGAGTATCTTTCGTACCATATGATGCTGAGGATATAGA 585  
 Db 535 TGTATGAAACAGCAAGCAAAAAATATACCTCTCTCTCTTACATATGTCGAGGTATTCGT 594  
 Qy 586 GCAGATCTTATCACTATTTTAAAGACCTCAATCTAAAATTTGCTTACCAAGGAAAAATA 645  
 Db 595 ACTGATTTAATTCACATGTTTGAAGTACACATCTTAAATTTCTTATCAAGGAAAGCTA 654  
 Qy 646 GGTATTAGTTACCTCATCACACCAAGAGTCTCTGCAATTTATGTTGGTATGATGCTACCTAC 705

```

Db      655 GGGTGGCTACTTCGTAAAGTCAGAGTCTTCGGTTCCTTTGGTATATATTTTCATAAA 714
Qy      706 GTTATTTGTAATAAATTTGAGAGATACCTGTAATAACCTCTGTAGTATTAATAATGATGCT 765
Db      715 ATTATAAATAAAGTATTAATAAATGTTCCAGCCATGGTACCTATTAACTCAGACGAGATA 774
Qy      766 CCTCAACACATCTGCTTCAGTAACTCTTGACGTTGGATATCTTTGGCGGAGAAATTTGGA 825
Db      775 GTAGACCACACAGTTTGCACACAGTAACATTAATGTATGCTACTTTTGGATTAGAAGCTTGA 834
Qy      826 ATGAGTTTCACCTTC 840
Db      835 TGTAGTTCAACTTC 849

RESULT 36
AAX34763
ID AAX34763 standard; DNA; 852 BP.
AC AAX34763;
XX
DT 05-JUL-1999 (first entry)
DE DNA encoding P30-3 protein.
XX
KW Outer membrane protein; OMP; Ehrlichia chafeensis; E. canis; P30;
  detection; dog; ss.
XX
OS Ehrlichia canis.
XX
PN WO9913720-A1.
XX
PD 25-MAR-1999.
XX
PF 18-SEP-1998; 98WO-US19600.
XX
PR 19-SEP-1997; 97US-0059353.
XX
PA (OHIS ) UNIV OHIO STATE.
XX
PI Ohashi N, Rikihisa Y;
XX
DR WPI: 1999-254290/21.
DR P-PSDB; AAY06963.
XX
PT Novel outer membrane proteins from Ehrlichia chafeensis and
  Ehrlichia canis
XX
PS Disclosure; Fig 23A; 55pp; English.
XX
CC The invention provides isolated outer membrane proteins (OMP) from
  Ehrlichia chafeensis and E. canis. The E. chafeensis proteins form part
  of the OMP family and consist of proteins OMP-1, -(B to Z) shown
  in AAY06943-958. The E. canis proteins form part of the P30 family and
  consist of proteins shown in AAY06959-970. The proteins and genes are
  used to detect E. chafeensis in patients and E. canis in dogs.
XX
SQ Sequence 852 BP; 295 A; 131 C; 152 G; 274 T; 0 other;

Query Match      21.4%; Score 179.8; DB 20; Length 852;
Best Local Similarity 53.5%; Pred. No. 5.7e-33;
Matches 457; Conservative 0; Mismatches 377; Indels 21; Gaps 3;

Qy      1 ATGAATTTATAAGAAAATTTCTAGTAAGACGGGTTAATCTCAATTAATGTCAATCTTTACCA 60
Db      1 ATGAATTTATAAGAAAATTTCTCAATAAGTGCATTGATATCATCCATATACCTCTACTCT 60
Qy      61 TATAGTCTTTTGGAGATCTCTGTAGGTTCAAGAACTAATGATGATAACAAGAGGCTTTCTAC 120
Db      61 AATGTCTCATACTCTAACCCAGTATATGTTAAACAGTATGTATGGTAAT-----TTTAC 114
Qy      121 ATTAGTGCAAGTACATCAAGTATATCACACTTTAGAAAATTTCTCTGCTGAAGAACT 180

```

```

Db      115 ATATCAGAAAGTACATGCCAAGTGTCTCTCATTTTGGAAATTTTTCAGCTGAAGAAGAG 174
Qy      181 CCTATTAAATGAAC---AAATTTCTCCTCACTAAAAAAGTTTTCGGACTAAAGAAAGATGGT 237
Db      175 AAAAAAAGACAACTGTAGTATATGGCTTAAAGAGAAAACCTGGCAGGAGATGCAATATCT 234
Qy      238 GATATAACAAAAAAGACGATTTTACAGAGATAGTCTCCAGGCATTTGATTTTCAAAAATAAC 297
Db      235 AGTCAAAAGTCAGATGATAAATTTTACCATTTCGAAATTAATCAATTCAGATATGCAAGCAAC 294
Qy      298 TTAATATCAGGATTTTTCAGGAAGTATTTGGTTACTCTATGGACGGACCAAGAAATAGAACTT 357
Db      295 AAGTTTTTAGGTTTTCAGTAGCTATTTGGTTACTTCGATAGGCAGTCCCAAGATAGAGTT 354
Qy      358 GAAGTCGATATCAACAAATTTTAATCCAAAAAACCAGTAACCAATGATATCTGA----- 410
Db      355 GAGATGTCTTATGAAGCATTTTGTGTAATAATCCAGGTGATAATTAACAAAACGGTCT 414
Qy      411 -----TAATGGTGAATACTATATAACATTTTGCATTAATCTCGTAAGATGCAATGAAGAT 465
Db      415 TACAGGTATTGTGCTTTATCTCATCAAGATGATCGGATGATGACACTAGTGCAACT 474
Qy      466 CAGCAATATGTAGTACTTAAAAATCAGCGCATAACTTTTATGTCTCATTTGATGGTTAATACT 525
Db      475 GACAAATTTGTATATTAATTAATGAAGGATTAATTAACATATCATTTATGACAAACATA 534
Qy      526 TGCTATGACATTAAGCTGAAGAGATATCTTTCTGATCCATATGATGTCAGGATATAGGA 585
Db      535 TGTATGAAACAGCAAGCAAAAATATACCTCTCTCTTACATATGTGCAGGATTTGGT 594
Qy      586 GCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAATA 645
Db      595 ACTGATTTAATTCACATGTTTGAACACTACACATCTTAAATTTCTTATCAAGGAAAGCTA 654
Qy      646 GGTATTAGTTACCTCTATCACACAGAGTCTCTGCATTTATTTGGTGAATATACCATGGC 705
Db      655 GGGTTGGCTACTTCGTAAGTGCAGAGTCTTCGGTCTTTTGGTATATATTTTCATAAA 714
Qy      706 GTTATTTGGTAATAAATTTGAGAAGATACCTGTATAAATCTCTGTAGTATTAATATGATGCT 765
Db      715 ATTATAAATAAAGTTTAAAAATGTTCCAGCCATGGTACCTATTAACTCAGACGAGATA 774
Qy      766 CCTCAACACCATCTGCTTCAGTAACCTTCAGCTTGGATACTTTGGCGGAGAAATTTGGA 825
Db      775 GTAGACCACAGTTTGCACACAGTAACTAATAATGTATGCTACTTTGGATTAGAAGCTTGA 834
Qy      826 ATGAGTTTCACCTTC 840
Db      835 TGTAGTTCAACTTC 849

RESULT 37
AAC68707
ID AAC68707 standard; DNA; 864 BP.
XX
AC AAC68707;
XX
DT 02-MAR-2001 (first entry)
DE Ehrlichia canis VSAL gene partial coding sequence.
XX
KW Ehrlichia canis; VSAL; variable surface antigen 1; MAP1;
  major antigenic protein 1; antirickettsial; vaccine; gene therapy;
  Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
  3gdorf3; ds.
XX
OS Ehrlichia canis.
XX
PN WO200065063-A2.
XX
PD 02-NOV-2000.
XX

```

Qy	508	TCATTGATGGTTAATACTTGTCTATGACATTACAGCTGAAGAGTATCTTTTGGTACCATAT	566
Db	534	TCATTTATGTTAAACGCATGTTATGACATAACAACCTGAAAAAATGCCCTTTTTCCACCTTAT	593
Qy	568	GCATGTGCAGGTATAGCAGAGACTTATCACTATTTTTTAAGACCTCAATCTRAAAATTT	627
Db	594	ATATGTGCAGGTATTTGGTACTCATCTCATATCTTATGTTTGAGACACACAAAACAATA	653
Qy	628	GCTTACCAAGGAAAAATAGGTATTAGTTTACCCCTATCACACCGAGAAGTCTCTGCAATTTAT	687
Db	654	TCTTTATCAAGGAAAGTTAGGTTTAAACTATACTATAAACTCAAGAGTTTCTGTTTTTGCA	713
Qy	688	GGTGGATACCAATGGCGGTTATTGGTAAATAAAATTTGAGAAGATACCTGTAATAACTCCT	747
Db	714	GGTGGGCACITTTCAATAGGTAATAGGTAATGAATTTAAAGGTATTCCTACTCTATTACCT	773
Qy	748	GTAGTATTAATGATGCTCTCTCAAACACACATCTGCTTCAGTAACTCTTTGACGTTGGATAC	807
Db	774	GATGGATCAACAATTAAAGTACAA--CAGTCTCGAACACAGTAAACATTAGATGTGTGCCAT	830
Qy	808	TTTGGCGGAGAAATTGGAATGAGGTTTCACTTT	839
Db	831	TTTCGGGTTAGAGATTGGAAGTAGATTTTCTT	862

  

RESULT 38	
AAV07180	AAV07180 standard; DNA; 1570 BP.
ID	XX
AC	AAV07180;
XX	XX
DT	14-SEP-1998 (first entry)
DE	• Ehrlichia canis VSA genomic locus.
XX	XX
KW	MAP1 homologue; variable surface antigen; VSA1, VSA2; rickettsia;
KW	DNA vaccine; ss.
XX	XX
OS	Ehrlichia canis.
XX	XX
Key	Location/Qualifiers
RBS	1..5
FT	/tag= a
FT	11...874
CDS	/tag= b
FT	/note= "VSA1 gene"
terminator	1015..1022
FT	/tag= c
terminator	1027..1034
FT	/tag= d
-35_signal	1081..1086
FT	/tag= e
exon	1101..1106
FT	/tag= f
RBS	1161..1165
FT	/tag= g
CDS	1171..1560
FT	/tag= h
FT	/note= "truncated VSA2 gene"
XX	XX
PN	WO9816554-A1.
XX	XX
PD	23-APR-1998.
XX	XX
Pf	17-OCT-1997; 97WO-US19044.
XX	XX
PR	17-OCT-1996; 96US-0733230.
XX	XX
PA	(UYFL ) UNIV FLORIDA.
XX	XX
PI	Barbet AF, Burrig MJ, Ganta RR, Mahan SM, McGuire TC;
PI	Nvika A, Rurangirwa FR;

XX WPI; 1998-251232/22.  
 DR P-PSDB; AAW51096-97.  
 XX  
 PT Composition containing nucleic acid encoding rickettsial antigen -  
 XX useful for, e.g. stimulating protective immune response in humans or  
 PT animals  
 XX  
 PS Claim 4; Fig 2C; 39pp; English.  
 XX  
 CC This is the DNA sequence of a 1.5 genomic locus of Ehrlichia canis  
 CC that was obtained using a PCR cloning strategy based on identifying  
 CC genes homologous to the major antigenic protein MAP1 (see AAW51088)  
 CC of Cowdria ruminantium. It includes 2 very similar but  
 CC non-identical open reading frames (ORFs), of which ORF2 is a  
 CC partial gene. Due to their similarity to MAP1 surface antigen  
 CC genes of C. ruminantium, the E. canis ORFs are designated variable  
 CC surface antigen (VSA) genes 1-2. A claimed composition comprises a  
 CC nucleic acid (see AAV07176-82) encoding a polypeptide (see AAW51088-99)  
 CC that elicits a protective immune response against a rickettsial  
 CC pathogen. The nucleic acid is used, in human or veterinary  
 CC medicine, in vaccines to protect against Rickettsia, Ehrlichia,  
 CC Anaplasma and Cowdria species. The nucleic acids are also useful  
 CC as probes to identify related sequences, e.g. for identification of  
 CC organisms and for diagnosing infection. Use of nucleic acid  
 CC vaccines avoids the problem of protein purification associated with  
 CC protein-based vaccines. The nucleic acid does not replicate in the  
 CC host but remains episomal and capable of expressing polypeptide for  
 CC at least 19 mth.  
 XX  
 SQ Sequence 1570 BP; 565 A; 223 C; 251 G; 531 T; 0 other;

Query Match 21.0%; Score 176.8; DB 19; Length 1570;  
 Best Local Similarity 55.0%; Pred. No. 3.3e-32;  
 Matches 480; Conservative 0; Mismatches 347; Indels 45; Gaps 5;

QY 1 ATGATTTATAGAAATCTAGTAAAGCGGTTAAATCTCAATTAATGCTCAATCTTACCA 60  
 DB 11 ATGAATATATAAAACATTTTACAGTAAGTCAATGATAGTATTAATTAATCTTCAACAT 70  
 QY 61 TATCAGTCTTTTGCAGATCTCTAGGTTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120  
 DB 71 TTTATACCTTTTATAGTCCAGCAGCTGCGAGTACAAATTCACAA-----CTTCTAC 121  
 QY 121 ATTAGTCAAGTACAATCCAAGTATATACACTTTAGAAAATCTCTGCT---GAAGAA 177  
 DB 122 ATTAGTGAATAATATATGCAACAGCGTCACATTTTGGAAATTTTTCAGCTAAAGAGAA 181  
 QY 178 ACTCTTATTAATGGAACAAATTTCTCTCAATAAAAGTTTTCGGACTAAAGAAAGATGCT 237  
 DB 182 CAAAGTTTCTAAGGTATTAAGTTGGTTAGTATCAACGNTATCAATATTAATAAAC 241  
 QY 238 GATATAACAAAAAGACGATTTTACAAGAGTAGCTCCAGGCAATTTGATTTTCAAAATAAC 297  
 DB 242 AATAATGATACAGCAAGAGTCTTAAGGTTCAAAATTTATTTAATAACAAATAAC 301  
 QY 298 TTAATATCAGGATTTTCAGGAAGTATTTGTTACTCTATGAGCGGACCAAGAAATAGACTT 357  
 DB 302 CCATTTCTAGGATTTTCAGGAGCTATTTGTTATTTCAATAGGCAATTTCAAGAAATAGACTA 361  
 QY 358 GAAGCTGCATATCAACAAATTTAATCCMAAAACACGATAACAAATGATGATGATGCT 417  
 DB 362 GAAGTATCAGTAAATATTTGATATCTAAACCCAGGAAACAAATTTAATAATGACTCT 421  
 QY 418 GAATATCTATAAACATTT-----TTGATTTATCTCTGTA 449  
 DB 422 CACAATATTTGCGCTTTATCTCATGGAAGTCACATATGAGTGAATGATGCGAGAT 481  
 QY 450 AGATGCAATGGAAGA--TCAGCAATATGATGATCTTAAATAATGACGGCAATACCTTTATG 507  
 DB 482 TGGTACACTGCAAAACACTGATAAGTTTGTACTTCTGAAATAATGAAGGTTTACTTTCAGCTC 541  
 QY 508 TCAATTGATGTTTAATTAATCTTGCTATGACATTTACAGCTGAAGGAGTATCTTTTCGTACCAT 567

DB 542 TCATTATGTTAAACGATGTTATGATACATACACTGAAAAATGCTTTTTCACCTTAT 601  
 QY 568 GCATGTCAGGTATAGGAGCAGATCTTATCATCTATTTTAAAGACCTCAATCTTAAATTT 627  
 DB 602 ATATGTCAGGTATTTGGTACTGATCTCATATCTATGTTTGGAGACACACAAAAAATA 661  
 QY 628 GCTTACCAGGAAAAATAGTATTAGTTACCTTATCACCCATCACACGAGAGTCTCTGCATTTAT 687  
 DB 662 TCTTATCAAGGAAAGTTAGGTTTAAACTATATAAATACTCAAGAGTTTCTGTTTTGCA 721  
 QY 688 GGTGATCTACTACCATGGCGTTATTGTTAAATAAATTTGAGAAGATACCTGTAATAACTCCT 747  
 DB 722 GGTGGGCACTTTCATAGGTAATAGGTATGAATTTAAAGGTATTCCTACTCTATTACCT 781  
 QY 748 GTAGTATTAAATGATGCTCTCAACACACATCTCTTCAGTAACTCTTCAGCTTGGATAC 807  
 DB 782 GATGATCAACACATTAAGATACAA---CAGTCTGCAACAGTAACATTAGATGTGTGCCAT 838  
 QY 808 TTTGCGGAGAAATGGAATGAGGTTTCACTT 839  
 DB 839 TTCGGGTAGAGATTGGAAGTAGATTTTCTT 870

RESULT 39  
 AAC68717  
 ID AAC68717 standard; DNA; 1570 BP.  
 XX AC AAC68717;  
 XX  
 DT 02-MAR-2001 (first entry)  
 XX  
 DE Ehrlichia canis 28 kDa gene locus.  
 XX  
 KW • Ehrlichia canis; MAP1; major antigenic protein 1; antirickettsial;  
 KW vaccine; gene therapy; Rickettsia; Cowdria; Anaplasma; map2; ihworf3;  
 KW 4hworf1; 18hworf1; 3gdorf3; ds.  
 XX  
 OS Ehrlichia canis.  
 XX  
 PN WO200065063-A2.  
 XX  
 PD 02-NOV-2000.  
 XX  
 PF 21-APR-2000; 2000WO-US10886.  
 XX  
 PR 22-APR-1999; 99US-0130725.  
 XX  
 PA (UYFL) UNIV FLORIDA.  
 XX  
 PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;  
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Allenman AR;  
 XX  
 XX WPI; 2000-679675/66.  
 XX  
 PT New polynucleotides useful as DNA vaccines for conferring immunity to  
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,  
 PT comprises major antigenic protein 1 or 2 Gene of rickettsial pathogens  
 PT  
 XX  
 PS Example 2; Fig 2C; 63pp; English.  
 XX  
 CC The present sequence is given in a specification relating to nucleic  
 CC acid vaccines which may be used to protect animals or humans against  
 CC rickettsial diseases caused by a organisms of Rickettsia sp.,  
 CC Ehrlichia sp., Anaplasma sp. or Cowdria sp. The vaccine elicits an  
 CC immune response protective against the rickettsial pathogen. The  
 CC vaccines comprises the major antigenic protein 1 (MAP1) or major  
 CC antigenic protein 2 (MAP2) of rickettsial pathogens. The nucleic acid  
 CC vaccines can be driven by the human cytomegalovirus (HCMV)  
 CC enhancer-promoter. Cowdria ruminantium genes designated map 2, ihworf3,  
 CC 4hworf1, 18hworf1 and 3gdorf3 may be used in therapeutic and diagnostic  
 CC applications. The polypeptides are useful for detecting antibodies

associated with infection by a rickettsial pathogen whilst the  
 CC polynucleotides may be used to detect the presence of rickettsial  
 CC nucleic acids.  
 XX  
 SQ Sequence 1570 BP; 565 A; 224 C; 251 G; 530 T; 0 other;  
 Query Match 21.0%; Score 176.8; DB 21; Length 1570;  
 Best Local Similarity 55.0%; Pred. No. 3.3e-32;  
 Matches 480; Conservative 0; Mismatches 347; Indels 45; Gaps 5

QY	1	ATGAATTTATAGAAAAATTTCTAGTAAAGACGCCGTTAAATCTCATTAATCTCAATCTTACCA	60
DB	11	ATGAATATATAAAAAATCTTTACAGTAACGTCATTAGTATTATTAACTTCTTTACACAT	70
QY	61	TATCAGTCTTTTGCAGATCCCTGTAGGTTCAAGAACTAATGATATAACAAGAGGCTTCTAC	120
DB	71	TTTATAGCTTTTTTATAGTCCAGCAGCTGCCAGTACAATTCAAA-----CTTCTAC	121
QY	121	ATTAGTCCAAAGTAGCAATCCAAGTATATCACACTTTAGAAAAATCTCTGCT---GAAGAA	177
DB	122	ATTAGTGGAAAAATATATGCCAAACAGCTCACTTTTGGATTTTTTTCAGCTTAAGAAGAA	181
QY	178	ACTCCTATTAAATGGAAACAAATTCCTCTCACTAAAAAAGTTTTTCGGACTAAAGAAAGATGGT	237
DB	182	CAAAAGTTTTACTAAGGTATTAGTTGGGTAGATCAACGATTATCACATAATATTATAAAC	241
QY	238	GATATAACAAAAAAGACGATTTTACAAGAGTAGCTCAGGCATTGATTTTCAAAATTAAC	297
DB	242	AAATAATGATACAGCAAAAGAGCTTTAAGGTTCCAAAATTTATTCACTTAAATACAAAAATTAAC	301
QY	298	TTAATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGGACGGACCAAGAAATAGAACTT	357
DB	302	CCATTTCTAGGATTTGCAGGAGCTATTGGTTATTCAATAGCAATTCAGAAATAGAACTA	361
QY	358	GAGGCTGCATATCAACAATTTTAAATCCAAAAACACCGAATAACAATGATCTGATAATGGT	417
DB	362	GAAGTATCACATGAAATATTGTGACTAAAAAACCCAGGAAACAAATTTTAAATGACTCT	421
QY	418	GAATACTATAAACATT-----TTGCATTATCTCGTAA	449
DB	422	CACAAATTTGGCGCTTTATCTCTGGAAGTCACATATGCAGTGATGGAATATGCGGAGAT	481
QY	450	AGATGCAATGGAAGA--TCAGCAATATGTAGTACTTAAAAATGACGGCAATACCTTTTATG	507
DB	482	TGGTACACTGCAAAAACTGATAAGTTGTACTCTTGAAAAATGAAGTTTACTTGACGTC	541
QY	508	TCATTGATGGTTAATCTTCTGCTATGACATTTACAGCTGAAGAGGATCTCTTCGTAACCAT	567
DB	542	TCATTTATGTTAAACCGCATGTTATGACATAACAACTGAAAAAATGCGCTTTTTCACCTTAT	601
QY	568	GCATGTCAGGTATAGCAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAAAATTT	627
DB	602	ATATGTCAGGTATTGTAATCTGATCTCATATCTATGTTTGAGACACACAAAAACAATA	661
QY	628	GCTTTACCAAGAAAAATAGGTATTAGTTACCTATCACACAGAAAGTCTCTGCATTTATT	687
DB	662	TCATTATCAAGGAAGTTAGGTTTAAACTATATACTATAACTCAAGAGTTTCTGTTTTTGCA	721
QY	688	GGTGGTACTACCATGCGGTTATTGGTTAATAAATTTGAGAAGATACCTGTGTAATACCTCT	747
DB	722	GGTGGGCACCTTTCAATAGGTAATAGGTAAAGAAATTTAAAGGATTTCCCTACTTATACCT	781
QY	748	GTAGTATTAAATGATGCTCTCTCAAAACACACATCTGCTTCAGTAACTCTTGACGTTGGATAC	807
DB	782	GATGGATCAAAACATTAAGTACAA--CAGTCTGCAACAGATAACATTAGATGTGTGCCAT	838
QY	808	TTTGGCGGGAATTTGGAATGAGGTTCACTTT	839
DB	839	TTCCGGTTAGAGATTTGGAAGTAGATTTTCTTT	870

RESULT 40  
AAS07583

ID	XX	AA007583 standard; DNA; 1570 BP.
XX	AC	AA007583;
XX	DT	23-OCT-2001 (first entry)
XX	DE	DNA encoding variable surface antigens 1-2 (VSA1-2) from <i>E. canis</i> .
XX	DE	Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
XX	KW	infection; heartwater; diagnostic; variable surface antigen; VSA; ds.
XX	OS	<i>Ehrlichia canis</i> .
XX	XX	
XX	Key	Location/Qualifiers
XX	RBS	1..5
XX	FT	/*tag= a
XX	FT	/note= "Ribosome binding site of VSA1 gene"
XX	FT	11..874
XX	FT	/*tag= b
XX	FT	/product= "Variable surface antigen 1 (VSA1)"
XX	FT	1015..1020
XX	FT	/*tag= c
XX	FT	/note= "Transcription terminator of VSA1 gene"
XX	FT	1027..1034
XX	FT	/*tag= d
XX	FT	/note= "Transcription terminator of VSA1 gene"
XX	FT	1081..1086
XX	FT	/*tag= e
XX	FT	/note= "-35 region of VSA2 gene"
XX	FT	1101..1106
XX	FT	/*tag= f
XX	FT	/note= "-10 region of VSA2 gene"
XX	FT	1151..1155
XX	FT	/*tag= g
XX	FT	/note= "Ribosome binding site of VSA2 gene"
XX	FT	1161..1169
XX	FT	/*tag= h
XX	FT	/product= "Variable surface antigen 2 (VSA2)"
XX	FT	/partial
XX	FT	/note= "No stop codon"
XX	XX	
XX	US6251872-B1.	
XX	XX	
XX	26-JUN-2001.	
XX	XX	
XX	17-OCT-1997;	97US-0953326.
XX	XX	
XX	17-OCT-1996;	96US-0733230.
XX	XX	
XX	(UYFL ) UNIV FLORIDA.	
XX	XX	
XX	Barbet AF, Ganta RR, McGuire TC, Burrige MJ, Nyika A;	
XX	Rurangirwa FR, Mahan SW, Bowie MV, Alleman AR;	
XX	XX	
XX	WPI; 2001-424487/45.	
XX	DR	P-PSDB; AAU04200, AAU04201.
XX	XX	
XX	XX	
XX	PT	New MAP2 genes and polypeptides useful as vaccines for conferring
XX	PT	immunity to human and animal rickettsial diseases, e.g. heartwater, or
XX	PT	as molecular markers in nucleic acid analysis procedures -
XX	XX	
XX	XX	Example 3; Fig 2C; 30pp; English.
XX	XX	
XX	CC	The sequence represents the coding sequence of variable surface antigen
XX	CC	(VSA) gene locus encoding VSA1-2 proteins of <i>Ehrlichia canis</i> , which
XX	CC	have similarity to major antigen proteins (MAP). The MAP polynucleotides
XX	CC	and polypeptides are useful as vaccines for conferring immunity to
XX	CC	rickettsia infection, including <i>Cowdria ruminantium</i> causing heartwater.
XX	CC	The MAP polynucleotides may be used as molecular markers in nucleic acid
XX	CC	analysis procedures, and to produce the MAP polypeptides, which may
XX	CC	be used to raise antibodies that are reactive with the polypeptides.
XX	CC	The nucleic acids may further be used as probes to identify
XX	CC	complementary sequences within other nucleic acid molecules or genomes.





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 00:15:51 ; Search time 46 Seconds  
(without alignments)  
5600.184 Million cell updates/sec

Title: US-10-062-624-41  
Perfect score: 840  
Sequence: 1 atgaattataagaataattct.....ttggaatgaggttcaccttc 840

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTCUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	840	100.0	840	4	US-09-660-587-41
2	243.4	29.0	849	4	US-09-648-520E-48
3	238.4	28.4	1607	4	US-09-660-587-1
4	238.4	28.4	1607	4	US-09-261-358A-1
5	238.4	28.4	1607	4	US-09-648-520E-47
6	238.4	28.4	1607	4	US-09-201-458-1
7	234	27.9	840	4	US-09-660-587-5
8	234	27.9	840	4	US-09-261-358A-5
9	228.6	27.2	830	4	US-08-953-326-11
10	212.8	25.3	861	4	US-08-953-326-8
11	207.2	24.7	843	4	US-08-953-326-10
12	205	24.4	837	4	US-08-953-326-9
13	203.8	24.3	842	3	US-08-733-230-3
14	203.8	24.3	842	4	US-08-953-326-3
15	195	23.2	828	4	US-09-660-587-43
16	195	23.2	864	3	US-08-733-230-1
17	195	23.2	864	4	US-08-953-326-1
18	181.4	21.6	849	4	US-09-660-587-3
19	181.4	21.6	849	4	US-09-261-358A-3
20	176.8	21.0	864	4	US-08-953-326-12
21	147.2	17.5	813	4	US-09-660-587-45
22	110.8	13.2	879	4	US-09-660-587-39
23	91.4	10.9	399	4	US-08-953-326-13
24	91	10.8	849	3	US-08-733-230-5
25	91	10.8	849	4	US-08-953-326-5
26	74.2	8.8	1242	4	US-09-288-339-1
27	74.2	8.8	1278	4	US-09-288-339-3

28	74.2	8.8	1919	4	US-08-975-762-40	Sequence 40, Appl
29	74.2	8.8	1919	4	US-09-295-028-40	Sequence 40, Appl
30	74.2	8.8	1919	4	US-09-106-582-40	Sequence 40, Appl
31	72.6	8.6	1176	4	US-09-288-339-5	Sequence 5, Appl
32	72.6	8.6	1840	4	US-09-066-046-26	Sequence 26, Appl
33	72.6	8.6	1843	4	US-09-066-047-12	Sequence 12, Appl
34	71	8.5	3435	4	US-09-066-046-28	Sequence 28, Appl
35	71	8.5	3435	4	US-09-066-047-14	Sequence 14, Appl
36	60	7.1	3435	4	US-09-066-046-28	Sequence 28, Appl
37	60	7.1	3435	4	US-09-066-047-14	Sequence 14, Appl
38	54.6	6.5	837	4	US-09-288-339-7	Sequence 7, Appl
39	50	6.0	12124	1	US-08-181-271A-36	Sequence 36, Appl
40	50	6.0	12124	1	US-08-449-315-36	Sequence 36, Appl
41	50	6.0	12124	1	US-08-444-803-36	Sequence 36, Appl
42	50	6.0	12124	1	US-08-449-043-36	Sequence 36, Appl
43	50	6.0	12124	1	US-08-456-265A-36	Sequence 36, Appl
44	50	6.0	12124	1	US-08-455-416-36	Sequence 36, Appl
45	50	6.0	12124	1	US-08-455-244-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1  
US-09-660-587-41  
; Sequence 41, Application US/09660587  
; Patent No. 6392023  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152C1p2  
; CURRENT APPLICATION NUMBER: US/09/660,587  
; CURRENT FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 41  
; LENGTH: 840  
; TYPE: DNA  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: nucleic acid sequence of E. canis p28-2  
US-09-660-587-41

Query Match	100.0%;	Score 840;	DB 4;	Length 840;
Best Local Similarity	100.0%;	Pred. No. 2.7e-208;		
Matches 840;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	ATGAATTTATAAGAAAATTTCTAGTAAGAGCGGTTAATCTCATTATGTCATCTTACCA	60	
Db	1	ATGAATTTATAAGAAAATTTCTAGTAAGAGCGGTTAATCTCATTATGTCATCTTACCA	60	
Qy	61	TATCAGTCTTTTGCAGATCTCTGAGGTTCAAGAACTAATGATTAACAAGAGGCTTCTAC	120	
Db	61	TATCAGTCTTTTGCAGATCTCTGAGGTTCAAGAACTAATGATTAACAAGAGGCTTCTAC	120	
Qy	121	ATTAGTCAAAAGTACAAATCCAAAGTATATACACCTTTAGAAAAATTTCTCTGCTGAAGAACT	180	
Db	121	ATTAGTCAAAAGTACAAATCCAAAGTATATACACCTTTAGAAAAATTTCTCTGCTGAAGAACT	180	
Qy	181	CCATTATTAATGGAACAAATTTCTCAGTAAAAAGTTTTTCGAGCTTAAAGAGATGGTGAT	240	
Db	181	CCATTATTAATGGAACAAATTTCTCAGTAAAAAGTTTTTCGAGCTTAAAGAGATGGTGAT	240	
Qy	241	ATAACAAAAAAGACGATTTTACAAGAGTAGCTCCAGGCAATTTGATTTTCAAAATAACTTA	300	
Db	241	ATAACAAAAAAGACGATTTTACAAGAGTAGCTCCAGGCAATTTGATTTTCAAAATAACTTA	300	
Qy	301	ATATCAGGATTTTCAGGAAGTATTTGGTTACTCTATGGAACGACCAAGAAATAGAACTTGAA	360	
Db	301	ATATCAGGATTTTCAGGAAGTATTTGGTTACTCTATGGAACGACCAAGAAATAGAACTTGAA	360	





```
QY      839 TC 840
Db      978 TC 979

RESULT 5
US-09-648-520E-47
; Sequence 47, Application US/09648520E
; Patent No. 6432649
; GENERAL INFORMATION:
; APPLICANT: Stich, Roger W.
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Methods for Detecting Ehrlichia Canis and Ehrlichia Chaffeensis in
; TITLE OF INVENTION: Vertebrate and Invertebrate Hosts
; FILE REFERENCE: 22727/04069
; CURRENT APPLICATION NUMBER: US/09/648,520E
; CURRENT FILING DATE: 2000-08-25
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47
; LENGTH: 1607
; TYPE: DNA
; ORGANISM: JAKE strain, Ehrlichia canis p30
US-09-648-520E-47

Query Match      28.4%; Score 238.4; DB 4; Length 1607;
Best Local Similarity 57.7%; Pred. No. 1.1e-52;
Matches 486; Conservative 0; Mismatches 346; Indels 10; Gaps 3;

QY      1 ATGAATTATAAGAAATTCCTAGTAAGACGGTGAATCTCAATTAATGCTCAATCTTACCA 60
Db      146 ATGAATTGCAAAAAAATTCCTATAACAACTGCATTAAATCAATTAATGCTCAATCTTCA 205

QY      61 TATCAGTCTTTTGCAGATCCTGTAGTTCACAACTCAATGATAACAAAGAGCTTTCTAC 120
Db      206 AGCATATCTTTTCTGATACTATACAGATGTTACATGGTGGTAA-----CTTCTAT 259

QY      121 ATTAGTGCAAGTATCAATCAAGTATATCACATTTTAGAAAAATTCCTGTCTGAAGA--AA 178
Db      260 ATTAGTGCAAGTATGTACCAAGTCTCTCACATTTTGGTAGCTTCTCAGCTAAAGAAGAA 319

QY      179 CTCCTTAATTAATGGAACAAATTCCTCACTAAAAAGTTTTCGGACTAAAGAAAGATGGTG 238
Db      320 AGCAAAATCAACTGTGTGGAGTTTTCGGATTAACATGATTTGGGATGGAAGTCCCAATCT- 378

QY      419 AATACATAAACAATTTTGCATTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTAG 478
Db      558 ACAGGTACTCGCTCTATCTCATCACATCGGCAGCCATGGAAGCTGATAAATTTGTCT 617

QY      479 TACTTAAATAAGACGCATCAACTTTTATGCTATGATGTTTAACTTCTGCTATGACATTA 538
Db      618 TCTTAAAAAACAAGGGTTTAAATGACATATCACTTGCATAAATAATGCAATGCTTATGATATA 677

QY      539 CAGCTGAAGAGTATCTTTCGTACCATATGCAATGTCAGTATAGGAGAGATCTTATCA 598
Db      678 TAAATGACAAAGTACTGTTTCTCTTATATATGCGCAGGTATTTGTTACTGATTTGATTT 737

QY      599 CTATTTTAAAGACCTCAATCTAAATTTTCTTACCAAGGAAAAATAGGTATTTAGTTACC 658
Db      738 CTATGTTTGAAGCTACAAGTCTTAAATTTCTCTACCAAGGAAAAATCTGGGCATTTAGTTACT 797

QY      559 CTATCACACCAGAAAGTCTCTGCATTATTTGGTGATACTACCATGGCGTTATTGTTAATA 718
Db      798 CTATTAAATCGGAAACCTCTGTTTTCATCGGTGGGCATTTCCACAGGATCATAGTTAATG 857

QY      719 AATTTGAGAGATACCTGTATAAATACTCTGTAGTATTAAATGATGCTCCTCAAAACCAAT 778
Db      858 AGTTTAGAGATATTCCTGCAATAGTACTAGTAACTCACTACATAAGTGGACCAAT 917

QY      779 CTGCTTCAGTAACTCTTTGACGTTGATCTTTGGCGGAGAAATTTGGAATAGGTTACCT 838
Db      918 TTGCAACAGATAACACTAAATGTGTGTCTGTTGTTAGAACTTTGGAGGAAGATTTAACT 977

QY      839 TC 840
Db      978 TC 979

RESULT 6
US-09-201-458-1
; Sequence 1, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; TITLE OF INVENTION: canis and Uses Thereof
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 1
; LENGTH: 1607
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: nucleic acid sequence of a gene encoding a 30 kDa
; OTHER INFORMATION: Immunoreactive protein of Ehrlichia canis
US-09-201-458-1

Query Match      28.4%; Score 238.4; DB 4; Length 1607;
Best Local Similarity 57.7%; Pred. No. 1.1e-52;
Matches 486; Conservative 0; Mismatches 346; Indels 10; Gaps 3;

QY      1 ATGAATTATAAGAAATTCCTAGTAAGACGGTGAATCTCAATTAATGCTCAATCTTACCA 60
Db      146 ATGAATTGCAAAAAAATTCCTATAACAACTGCATTAAATCAATTAATGCTCAATCTTCA 205

QY      61 TATCAGTCTTTTGCAGATCCTGTAGTTCACAACTCAATGATAACAAAGAGCTTTCTAC 120
Db      206 AGCATATCTTTTCTGATACTATACAGATGTTACATGGTGGTAA-----CTTCTAT 259

QY      121 ATTAGTGCAAGTATCAATCAAGTATATCACATTTTAGAAAAATTCCTGTCTGAAGA--AA 178
Db      260 ATTAGTGCAAGTATGTACCAAGTCTCTCACATTTTGGTAGCTTCTCAGCTAAAGAAGAA 319

QY      179 CTCCTTAATTAATGGAACAAATTCCTCACTAAAAAGTTTTCGGACTAAAGAAAGATGGTG 238
Db      320 AGCAAAATCAACTGTGTGGAGTTTTCGGATTAACATGATTTGGGATGGAAGTCCCAATCT- 378

QY      419 AATACATAAACAATTTTGCATTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTAG 478
Db      558 ACAGGTACTCGCTCTATCTCATCACATCGGCAGCCATGGAAGCTGATAAATTTGTCT 617

QY      479 TACTTAAATAAGACGCATCAACTTTTATGCTATGATGTTTAACTTCTGCTATGACATTA 538
Db      618 TCTTAAAAAACAAGGGTTTAAATGACATATCACTTGCATAAATAATGCAATGCTTATGATATA 677

QY      539 CAGCTGAAGAGTATCTTTCGTACCATATGCAATGTCAGTATAGGAGAGATCTTATCA 598
Db      678 TAAATGACAAAGTACTGTTTCTCTTATATATGCGCAGGTATTTGTTACTGATTTGATTT 737

QY      599 CTATTTTAAAGACCTCAATCTAAATTTTCTTACCAAGGAAAAATAGGTATTTAGTTACC 658
Db      738 CTATGTTTGAAGCTACAAGTCTTAAATTTCTCTACCAAGGAAAAATCTGGGCATTTAGTTACT 797
```

QY 419 AATCTATAAACAATTTTGCATTAATCTCGTAAGATGCAATGGAAGATCAGCAATATGTAG 478  
Db |||||  
QY 558 ACAGTACTGGCTCTATCTCATCACATCGGCGAGCCATGGAAGCTGATAAATTTGTCT 617  
Db |||||  
QY 479 TACTTAAAAATGACGGCATAACTTTTATGTCAATGATGGTTAAATCTTGTCTATGACATTA 538  
Db |||||  
QY 618 TCTTAAAAAAGCAGAGGTTAATGACATATCACTTGCAATTAATGATGATGATGATTA 677  
Db |||||  
QY 539 CAGCTGAAGAGTATCTTTTCGTACATATGATGTCAGAGGTATAGGACAGATCTTATCA 598  
Db |||||  
QY 678 TAAATGACAAAGTACTCTTCTCTTATATATGCGCAGGATTTGGTACTGATTTGATT 737  
Db |||||  
QY 599 CTATTTTAAAGACCTCAATCTAAATTTGCTTACCAGGAAATAGGATTTAGTTACC 658  
Db |||||  
QY 738 CTATGTTTGAAGCTACAAAGTCTTAAATTTCTTACCAGGAAATAGGATTTAGTTACT 797  
Db |||||  
QY 659 CTATCACACAGAGTCTCTGCAATTTATGTTGGATATCTACATGGCGTTATTTGTTATA 718  
Db |||||  
QY 798 CTATTAATCCGMAACCTCTGTTTTCATCGTGGGCAATTTCCACAGGATCATAGTTAATG 857  
Db |||||  
QY 719 AATTTGAGAAGTACCTGTAATAACTCTCTGATGATTTAAATGATCTCTCTCAAAACCAT 778  
Db |||||  
QY 858 AGTTTAGAGATATCTCTGCAATAGTACCTAGTAACTCAACTACAATAAGTGGACCAAT 917  
Db |||||  
QY 779 CTGCTTCAGTAACCTTTCAGCTTGATCTTGGATCTTTGGCGGAAATTTGGAATGAGTTTCACT 838  
Db |||||  
QY 918 TTGCAACAGTAACTAAATGTGTCTACCTTTGGTTTGAACCTTGGAGGAAGATTTTAACT 977  
Db |||||  
QY 839 TC 840  
Db |||||  
QY 978 TC 979

## RESULT 7

US-09-587-5  
; Sequence 5, Application US/09660587  
; Patent No. 6392023  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: US/09/660,587  
; CURRENT FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 5  
; LENGTH: 840  
; TYPE: DNA  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; OTHER INFORMATION: nucleic acid sequence of p28-6  
US-09-660-587-5

Query Match 27.9%; Score 234; DB 4; Length 840;  
Best Local Similarity 59.0%; Pred. No. 1.2e-51;  
Matches 503; Conservative 0; Mismatches 325; Indels 24; Gaps 5;  
QY 1 ATGAATTATAAGAAAATTTCTAGTAAGAGCGGTTAATCTCATTAATGCTCAATCTTACCA 60  
Db |||||  
QY 1 ATGAATTGCAAAAAATTTCTTAAACAACACTGCAATTAATGCTCAATTAATGCTACTATGCTCCA 60  
Db |||||  
QY 61 TATCAGTCTTTGAGATCTCTGATGTTCAAGAAGTAAATGATATAACAAGAGGCTTCTAC 120  
Db |||||  
QY 61 AGCATATCTTTTCTGTACTATATAAG-----ACGATAACACTGGTAGCTTCTAC 111  
Db |||||  
QY 121 ATTAGTCAAAAGTACAATCCAAAGTATATCACTTTTGAATAATTTCTCTGCTGAGA--AA 178  
Db |||||  
QY 112 ATCAGTGAAAATATGTACCAAGTGTTCACATTTTGGTGTTCACATTTTGGTGTTCACAGCTAAAGAA 171  
Db |||||

QY 179 CTCCTATTAATGGAACAAATTTCTCTCACTAAAAAGTTTTCGGACTAAGAAAAGATGGTG 238  
Db |||||  
QY 172 AGAAACTCAACTGTTGGAGTTTTTGGATTAAAAATGATTTGGAATGGAGGTACAATACTCT 231  
Db |||||  
QY 239 ATATAACAAAAAAGACGATTTTACAGAGTAGCTCCAGGCAATTTTTC-AAAAATAC 297  
Db |||||  
QY 232 AACTCTTCTCCAGAAAATATATTTCCAGTTTCAAAATTTATTCGTTTAAATACGAAAAACAC 291  
Db |||||  
QY 298 TTAATATCAGGATTTTTCAGGAAGTATTTGTTTACTCTATGACGACCAAGCAATAGAACTT 357  
Db |||||  
QY 292 CCATTTCTAGGTTTTCAGGAGCTATTTGTTTATTCATTTGTTGGTGGCCCAAGAAATAGAACTT 351  
Db |||||  
QY 358 GAAGCTGCATATCAAAATTTTAAATCCAAAAAACACCGATACAAATGATCTGATAATGGT 417  
Db |||||  
QY 352 GAAGTCTGTACGAGACATTCGATGTGAAAAATCAGAAACAATAATTAAGAACGGCGCA 411  
Db |||||  
QY 418 GAATACTATAACATTTTTCGATTTATCTCTGTAAGATGCAATG-----GAAGATCAG 468  
Db |||||  
QY 412 CACAGATACGTGCTTTTATCTCATCATGTTTCAGCAACNAGCATGTCTCCGCAAGTAAAC 471  
Db |||||  
QY 469 CAATATGTAGTACTTAAAAATGACGCAATAACTTTTATGTCAATGATGTTTAAATACTTGC 528  
Db |||||  
QY 472 AATTTGTTTCTTAAAAAATGAAGGTTAATTCGACTTATCATTTTATGATAAATGCAATGC 531  
Db |||||  
QY 529 TATGACATTAACAGCTGAAGAGTATCTTTCTGATACCATATGATGTCAGGTTATAGGAGCA 588  
Db |||||  
QY 532 TATGACATAATAATTGAAGGAATGCCITTTTTCACCTTATATTTTGTGCAAGGTGTTGGTACT 591  
Db |||||  
QY 589 GATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAATAGGT 648  
Db |||||  
QY 592 GATGTTGTTTCCATGTTTGAAGCTATAAATTTCTTAAATTTTCTTACCAAGGAAAACTAGGA 651  
Db |||||  
QY 649 ATTAGTTTACCCTATCACACAGAGTCTCTGCAATTTTATGTTGGATGATACTACCATGGCGTT 708  
Db |||||  
QY 652 TTAGTTTATAGTATAAGTTTCAAGAGCTCTGTTTATCGTGGACACTTTCACAGATC 711  
Db |||||  
QY 709 ATTGGTAAATAAATTGAGAAGATACCTGTATAACTCTCTGTAGTATTAATGATGCTCTCT 768  
Db |||||  
QY 712 ATAGTAAATGAATTTAGAGACATCCCTGCTATGTTTCTTAGTGGATCAA---ATCTTCCA 768  
Db |||||  
QY 769 CAACACACATCTGCTTCAGTAACTTTGACGTTGGATACCTTTGGCGGAGAAATTTGGAATG 828  
Db |||||  
QY 769 GAAACCAATTTGCAATAGTAAACACTAAATGTGTCTACCTTTGGCATAGAACTTTGGAGGA 828  
Db |||||  
QY 829 AGGTTTCACTTC 840  
Db |||||  
QY 829 AGATTTAACTTC 840  
Db |||||

## RESULT 8

US-09-261-358A-5  
; Sequence 5, Application US/09261358A  
; Patent No. 6403780  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP  
; CURRENT APPLICATION NUMBER: US/09/261,358A  
; CURRENT FILING DATE: 1999-03-03  
; PRIOR APPLICATION NUMBER: 09/201,458  
; PRIOR FILING DATE: 1998-11-30  
; NUMBER OF SEQ ID NOS: 33  
; SEQ ID NO 5  
; LENGTH: 840  
; TYPE: DNA  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; OTHER INFORMATION: nucleic acid sequence of ECa28SA3

US-09-261-358A-5

Query Match 27.9%; Score 234; DB 4; Length 840;  
Best Local Similarity 59.0%; Pred. No. 1.2e-51;  
Matches 503; Conservative 0; Mismatches 325; Indels 24; Gaps 5;

Qy 1 ATGATTTATAGAAAATCTAGTAGAGCGGTAAATCTCAATTAATGCTCAATCTTACCA 60  
Db 1 ATGAAATGCAAAAATCTTATACAACTGCAATTAATGCTCAATGCTCTCA 60

Qy 61 TATCAGTCTTTGCGAGCTCTAGGTTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120  
Db 61 AGCATATCTTTTGTGATCTATACAAG-----ACGATAACACTGGTAGCTTCTAC 111

Qy 121 ATTAGTGCAGAGTACAATCAAGTATATACACTTTAGAAAATCTCTGCTGAAGA--AA 178  
Db 112 ATCAGTGAAATATATGACCAAGTCTTACACTTTTGGTGTCTCTCAGCTAAAGAGAA 171

Qy 179 CTCCTTAAATGGAACAATCTCTCACTAAAAGTCTTCCGACTAAAGAAAGTGGT 238  
Db 172 AGAAACTCAACTGTTGGAGTTTGGATTAAAACATGATTGGAATGGAGGTACAATATCT 231

Qy 239 ATATAACAAAAAGACGATTTTACAAGTAGTCTCCAGGCAATTCATTTTC-AAAAATAC 297  
Db 232 AACTCTTCTCCAGAAAATATATTCACAGTTCAAATTTATTCGTTTAAATACGAAACAC 291

Qy 298 TTAATATCAGGATTTTTCAGGAAGTATTTGGTTACTCTATGGACGACCAAGAAATAGAACTT 357  
Db 292 CCATCTCTTAGGTTTTCAGGAGCTATTTGGTTATCAATGGGTGGGCCAAGAAATAGAACTT 351

Qy 358 GAAGCTGCATATCAACATTTTAAATCCAAAACACCGATAACAATGATCTGATATGGT 417  
Db 352 GAAGTTCTGTACGAGACATTCGATGTGAAAAATCAGAACAAATTTTAAAGAACGCGCA 411

Qy 418 GAATACTATAAACATTTTGCATTTATCTCGTAAAGATGCAATG-----GAAATCAG 468  
Db 412 CACAGTACTGTGCTTTTATCTCATATAGTTTCCAGCAACAGCATGTCCTCCGCAAGTAA 471

Qy 469 CAATATGTAGTACTTAAATGACGCGCATAACTTTTATGTCATTTGATGTTTAAATCTTGC 528  
Db 472 AAATTTGTTTCTTAAAAAATGAAGGGTAAATGACTTATCATTTTATGATAAATGCATGC 531

Qy 529 TATGACATACAGCTGAGGAGTATCTTTCGTAACCATATGATGTCAGGTATAGAGCA 588  
Db 532 TATGACATAAATTAAGAGAAATGCTTTTTCACCTTATTTTGTGCGAGGTGTTGCTACT 591

Qy 589 GATCTTATCACATTTTAAAGACCTCAATCTAAAATTTTGCTTACCAAGGAAAAATAGGT 648  
Db 592 GATGTTGTTTCCATGTTTGAAGCTATAAATCTTAAATTTCTTACCAAGGAAAACTAGGA 651

Qy 649 ATTAGTTACCTTATCACACAGAGTCTCTGCAATTTATTTGGTGGATATACCATGGCGTT 708  
Db 652 TTAGGTTATAGTATAAGTTTCAAGAGCCTCTGTTTTATCGTGGACACTTTTCACAGAGTC 711

Qy 709 ATTGGTAAATATTTTACAGAGATACCTGTAATAACTCTCGTAGTATTAATGATGCTCT 768  
Db 712 ATAGGTAATGAATTTAGACATCCCTGCTATGGTTCTTAGTGGATCAA---ATCTTCCA 768

Qy 769 CAAACACATCTGCTTTCAGTAACTCTTGACGTTGGATACCTTTGGCGGAGAAAATGGAATG 828  
Db 769 GAAACCAATTTGCAATAGTAACTAAATGTTGTCACCTTTGGCATAGAACTTGGAGGA 828

Qy 829 AGGTTCACTTC 840  
Db 829 AGATTTAACTTC 840

## RESULT 9

US-08-953-326-11  
; Sequence 11, Application US/08953326  
; Patent No. 6251872  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.

; APPLICANT: Ganta, Roman R.  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Nyika, Aceme  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Mahan, Suman M.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of  
; FILE OF INVENTION: Animals and Humans  
; FILE REFERENCE: UF-167C1  
; CURRENT APPLICATION NUMBER: US/08/953,326  
; EARLIER FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/953,326  
; EARLIER FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/733,230  
; EARLIER FILING DATE: 1996-10-17  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 830  
; TYPE: DNA  
; ORGANISM: Ehrlichia chaffeensis  
US-08-953-326-11

Query Match 27.2%; Score 228.6; DB 4; Length 830;  
Best Local Similarity 58.2%; Pred. No. 3e-50;  
Matches 487; Conservative 0; Mismatches 329; Indels 21; Gaps 4;

Qy 1 ATGAATTATAGAAAATTTCTAGTAAGAGCGGTTAAATCTCAATTAATGCTCAATCTTACCA 60  
Db 3 ATGAATTACAAAAAAGTTTTTCATAACAGTCAATGATATCAATTAATATCTTCTCTACCT 62

Qy 61 TATCAGTCTTTGCGAGTCTCTAGTTCAGAACTAATGATAACAAAGAGGCTTCTAC 120  
Db 63 GGAGTATCATTTTCCGACCCAGCAGGTAGTGTATTAAACGGTAAT-----TTCTAC 113

Qy 121 ATTAGTGCAGAAATACAATCCAAAGTATATCAACATTTAGAAAAATCTCTGCT---GAAGAA 177  
Db 114 ATCAGTGGAAAATACATGCGCAAGTCTTCGCAATTTTGGAGTATTTCTCTGCTAAGGAAGA 173

Qy 178 ACTCCTATTAATAGAAACAAATTTCTCTCACTAAAAGGTTTTCGAGCTAAAGAAAGATGGT 237  
Db 174 AGAAATAACAAGTTGGAGTGTTTGGACTGAAGCAAAAATTTGGGACGGAAGCGCAATATCC 233

Qy 238 GATATAACAAAAAGAGCATTTTACAAGAGTAGCTCCAGGCATTTGTTTCAAAATAAC 297  
Db 234 AACTCTCTCCAAACGATGATTCAGTCTCTCAAAATTTATTCATTTTAAATATGAAAAAAC 293

Qy 298 TTAATATCAGGATTTTTCAGGAAGTATTGGTTACTCTATGGACGGAACCAAGAAATAGAACTT 357  
Db 294 CCGTTTTTAGGTTTTCAGGAGCTATTGGTTACTCAATGGATGGTCCCAAGAAATAGAGCTT 353

Qy 358 GAAGCTGCATATCAACAAATTTTAAATCCAAAAACCGATAACAATGATGATCTGATAATGGT 417  
Db 354 GAAGTATCTTATGAACATTTTGATGTAAAAAATCAAGGTAAACAATTTAAGAAATGAAGCA 413

Qy 418 GAATACTATAACATTTTGCATTTATCTCGTAAAGATGCAATG-----GAAGATCAGCAA 471  
Db 414 CATAGATATTGTGCTCTATCCATAACTCAGCAGCAGACATGAGTAGTGCAGTAATTAAT 473

Qy 472 TATGTAGTACTTAAAAATGAGGCAATACTTTTATGTCAATGATGGTTTAAATCTTGTCTAT 531  
Db 474 TTTGCTTCTTAAAAAATGAAGGATTACTTGACATATCATTTATGCTCAAGCATGCTAT 533

Qy 532 GACATTAAGCTGAAGGAGTATCTTTGCTGTAACAATGATGTCAGGATATAGGAGCAGAT 591  
Db 534 GACGTAGTAGGCAAGGCAATCTTTTCTCTATATATATGCGCAGGTATCGGTACTGAT 593

Qy 592 CTTATCACTATTTTAAAGACTCAATCTAAAATTTGCTTAACTTAACTAGGTATT 651  
Db 594 TTAGTATCCATGTTTGAAGCTACAATCTTAAATTTCTTACCAAGGAAATGTTAGGTTTA 653

Qy 652 AGTTACCCCTATCACACCAGAAAGTCTCTGCAATTTTATTTGGTGATATCTACCATGGCGTATT 711

Db 654 AGTACTCTATAAGCCAGAACGCTTCTGTGTTTATTGTGGCAGCTTTCATAGGTAATA 713  
Qy 712 GGTATAAATTGAGAGATAC---CTGTAATAACTCTCTGTAGTANTTAATGATGCTCCT 768  
Db 714 GGAACGAATTTAGAGATATCTTACTATAATACCTACTGGATCAACACTTGCAGGAAA 773  
Qy 769 CAACACACATCTGCTTCACTAGTAACTCTTGACGTTGGATCTTTGGCGGAGAAATTTGGA 825  
Db 774 GGAACTACCTGCAATAGTAATACTGGATGTATGCCACTTTGGAAATAGAAATGGGA 830

RESULT 10  
US-08-953-326-8  
; Sequence 8, Application US/08953326  
; Patent No. 6251872  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Ganta, Roman R.  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Nyika, Aceme  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Mahan, Suman M.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of  
; TITLE OF INVENTION: Animals and Humans  
; FILE REFERENCE: UP-167C1  
; CURRENT APPLICATION NUMBER: US/08/953.326  
; EARLIER FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/953.326  
; EARLIER FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/733.230  
; EARLIER FILING DATE: 1996-10-17  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 861  
; TYPE: DNA  
; ORGANISM: Ehrlichia chaffeensis  
US-08-953-326-8

Query Match 25.3%; Score 212.8; DB 4; Length 861;  
Best Local Similarity 57.3%; Pred. No. 3.6e-46;  
Matches 492; Conservative 0; Mismatches 342; Indels 24; Gaps 5;

Qy 1 ATGAATTATAAGAAAATTTCTAGTAAGAGCGGTTAATCTCAATTAATGCTCAATCTTACCA 60  
Db 3 ATGAAGTCGCAAAAATTTTATAACCACTGCAATTAACATTAATGCTCTTCTTACCT 62  
Qy 61 TATCAGTCTTTGCGATCTGTTAGGTTCAAGAACTAATGATAACAAGAAAGGCTTTCTAC 120  
Db 63 GGAATATCACTTCTGATCCAGTACAG-----GATGACAACATTAAGTGGTAATTTCTAC 116  
Qy 121 ATTAGTCAAGAGTACAATCCAGTATATCACTTTAGAAAATTTCTGCT---GAGAA 177  
Db 117 ATCAGTGAAGTATATGCCAAGGCTTCGCAATTTGGAGTTTTTCTGCCAAGGAAGAA 176  
Qy 178 ACTCCTATTAATGAACAATAATCTCTCACTAAAAAGTTTTTCGCACTAAAGAAAGATGGT 237  
Db 177 AGAAATACAACAGTTGGAGTATTTGGAATAGACGAGATTTGGATAGATGTAATATCT 236  
Qy 238 GATATACAAAAAAGACGATTTTACAGAGTAGTCTCAGGCAATGATTTTCAAAAATAC 297  
Db 237 AGAACCACTTTAAGCGATATATTCACCGTTCCAAATTTATTCATTTAAGTATGAAAAAT 296  
Qy 298 TTAATATCAGATTTTCAGGAGTATTTGGTTACTCTTATGACGCGACCAAGATAGACTT 357  
Db 297 CTATTTTCAGATTTGCAAGAGCTATTGGCTACTCAATGGATGGCCCAAGAAATAGAGCTT 356  
Qy 358 GAAGTGCATATCAACAATTTAATCCAAAAACACCGATAACAATGATCTGATAATG--- 415  
Db 357 GAAGTATCTTATGAAGCATTCGATGTTTAAAAATCAAGGTACATTAATTAAGAACGAAGCA 416  
Qy 416 -GTGAATACTATAAACAATTTTGGCAATATCTCGTAAAGATGCAATGGAAGATCAGCAA--- 471

Db 417 CATAGATATTATGCTGTGTCCTCCATCTTCTCGGCACAGACACAGATAGATGGTGCAGGC 476  
Qy 472 -----TATGTAGTACTTAAAAATGACGGCATAACTTTTATGTCAATTGATGTTTAATACT 525  
Db 477 AGTCGGTCTGTCTTCTTAATAAATGAAGGACTCTTGATAAATCATTTATGCTGACCGCA 536  
Qy 526 TGCTATGACATTAACAGCTGAAGAGTATCTTTCGTACCATATGATGCTGACAGGATAGGA 585  
Db 537 TGTATGATGTAATAAGTGAAGGCATACCTTTTTCTCTTATATATATGTCAGGATTTGGT 596  
Qy 586 GCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAATA 645  
Db 597 ATTGAATTTAGTATCCATGTTTGAAGCTATAAATCCTAAAAATTTCTTATCAAGGAAAAATTA 656  
Qy 646 GGTATTAGTTACCTTATCACACACCAAGAGTCTCTGCATTTTATTTGGTGATCTACCATGGC 705  
Db 657 GGCTTAAGTTACCTTATAAGCCCAAGAGCTTCTGTGTTTATTTGGTGACATTTTCATNAAG 716  
Qy 706 GTTATTGGTAATAAATTTGAGAAGATACCTGTATAAATACTCCTGTAGTATTAATGATGCT 765  
Db 717 GTGATAGGAAACGAATTTAGAGATATTTCTTACTATGATACCTAGTGAATCAGCGCTTGCA 776  
Qy 766 CCTCA---AACCACATCTGCTTCAGTAACCTTTGACGTTGGATACTTTGGCGGAGAAATT 822  
Db 777 GGAAAAGGAAACTACCTGCAATAGTAACACCTGGACGTTTCTACTTTGGCATAGAACTT 836  
Qy 823 GGAATGAGGTTTCACCTTC 840  
Db 837 GGAGGAAGGTTTAACTTC 854

RESULT 11  
US-08-953-326-10  
; Sequence 10, Application US/08953326  
; Patent No. 6251872  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Ganta, Roman R.  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Nyika, Aceme  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Mahan, Suman M.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of  
; TITLE OF INVENTION: Animals and Humans  
; FILE REFERENCE: UP-167C1  
; CURRENT APPLICATION NUMBER: US/08/953.326  
; EARLIER FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/953.326  
; EARLIER FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/733.230  
; EARLIER FILING DATE: 1996-10-17  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 843  
; TYPE: DNA  
; ORGANISM: Ehrlichia chaffeensis  
US-08-953-326-10

Query Match 24.7%; Score 207.2; DB 4; Length 843;  
Best Local Similarity 56.2%; Pred. No. 1e-44;  
Matches 477; Conservative 0; Mismatches 353; Indels 18; Gaps 4;

Qy 1 ATGAATTATAAGAAAATTTCTAGTAAGAGCGGTTAATCTCAATTAATGCTCAATCTTACCA 60  
Db 3 ATGAATTTGCAAAAAATTTTATAACCACTACATTAAGTACGTTAAATGCTTCTTACCT 62  
Qy 61 TATCAGTCTTTGCGATCTCTGAGTCTTCAAGACTAATGATAACAAGAAAGGCTTCTAC 120  
Db 63 GGAATATCAATTTTCTGATGAGTACAGAACCAATGTTGGTGGTAAT-----TTCTAT 116





RESULT 13  
US-08-733-230-3  
; Sequence 3, Application US/08733230  
; Patent No. 6025338  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Ganta, Roman Reddy  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Nyika, Aceme  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Mahan, Suman M.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against  
; TITLE OF INVENTION: Rickettsial Diseases and Methods of Use  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/733,230  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Whitlock, Ted W.  
; REGISTRATION NUMBER: 36,965  
; REFERENCE/DOCKET NUMBER: UP-167  
; TELEPHONE: 352-375-8100  
; TELEFAX: 352-372-5800  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 842 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..840  
US-08-733-230-3

Query Match 24.3%; Score 203.8; DB 3; Length 842;  
Best Local Similarity 57.8%; Pred. No. 7.6e-44;  
Matches 424; Conservative 0; Mismatches 297; Indels 12; Gaps 3;  
QY 115 TTCTACATTAGTGCAAGTACCAATCCAGTATATACACTTTAGAAAATCTCTGCT--- 171  
DB 109 TTCTACATTAGTGCAAGTACCAATCCAGTATATACACTTTAGAAAATCTCTGCTAAG 168  
QY 172 GAAGAACTCTATTATATGGAACAATCTCTCACTAAAAAGTTTCGGACTAAAGAA 231  
DB 169 GAAGAAAGAAATACAACTGTTGGAGTGTTCAGTGAAGCAAAATTTGGACGGAAGCGCA 228  
QY 232 GATGCTGATATAACAAAAAGACCATTTTCAAGAGTAGTCCAGGCGATTGTTTCAAA 291  
DB 229 ATATCCAACTCTCCCAAGAGATGTTACTGTCTCAATATTATCAATTAATATGAA 288  
QY 292 AATAACTTAATACAGGATTTTCAGGAAGTATGTTACTCTATGACGCGACCAAGATA 351  
DB 289 AACAAACCGTTTTAGGTTTTGCAGGAGCTATTGTTACTCAATGATGTTGCCAAGATA 348  
QY 352 GAACTTGAGCTGATATCAACAATTTAATCCAAAAACCGGATAACAATGATGAT 411

DB 349 GAGCTTGAAGTATCTTATGAAAACATTTGATGTAATAAAATCAAGGTAAACAATTAAGAAT 408  
QY 412 AATGGTGAATATAAAACATTTTGCATTTATCTCGTAAAGATGCAATG-----GAAGAT 465  
DB 409 GAAGCACATAGATATTGTGCTCTATCCATAAATCTCAGCAGCAGATGAGTAGTCAAGT 468  
QY 466 CAGCAATATGATGATCTTTAAAAATGACGGCATAAATTTTATGTCATTGATGGTTAAATCT 525  
DB 469 AATAAATTTTGTCTTTCTAAAAAATGAAGGATTAATCTTGACATATCATTTTATGCTGAACGCA 528  
QY 526 TGCTATGACATTAACAGCTGGAAGGAGTATCTTTCTGACCATATGATGTCAGGATATAGGA 585  
DB 529 TGCTATGACGATGAGGAGGAGGATACCTTTTCTCTTATATATGCGCAGGATTCGGT 588  
QY 586 GCAGATCTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGAAAAATA 645  
DB 589 ACTGATTAGTATCCATGTTTGAAGCTACAAATCTTAAATTTCTTACCAAGGAAAGTTA 648  
QY 646 GGTATTAGTTACCTATACACACAGAGATCTCTGCAATTTATGTTGGATATCTACCATGGC 705  
DB 649 GGTAAAGCTACTCTATAAGCCCAAGAGCTTCTGTGTTTATTTGGTGGGCACTTTTCATAAG 708  
QY 706 GTTATTGTAATAAATTTGAGAAGTAC---CTGTAATACTCTCTAGTATTAATGAT 762  
DB 709 GTAATAGGGAACGAATTTAGAGATATTCCTACTATAATACCTACTGGAATCAACTTGCA 768  
QY 763 GCTCCTCAACACCATCTGCTTCAAGTAACTCTTGACGTTGGATATCTTTGGCGGAGAAAT 822  
DB 769 GGAAGAGGAACCTACCCTGCAATAGTAACTAGTATGATGATGATGATGATGATGATGATG 828  
QY 823 GGAATGAGGTTCA 835  
DB 829 GGAGGAAGGTTTA 841

## RESULT 14

US-08-953-326-3  
; Sequence 3, Application US/08953326  
; Patent No. 6251872  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Ganta, Roman R.  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Nyika, Aceme  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Mahan, Suman M.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of  
; TITLE OF INVENTION: Animals and Humans  
; FILE REFERENCE: UP-167C1  
; CURRENT APPLICATION NUMBER: US/08/953,326.  
; CURRENT FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/953,326  
; EARLIER FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/733,230  
; EARLIER FILING DATE: 1996-10-17  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 842  
; TYPE: DNA  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(840)  
US-08-953-326-3

Query Match 24.3%; Score 203.8; DB 4; Length 842;  
Best Local Similarity 57.8%; Pred. No. 7.6e-44;  
Matches 424; Conservative 0; Mismatches 297; Indels 12; Gaps 3;  
QY 115 TTCTACATTAGTGCAAGTACCAATCCAGTATATACACTTTAGAAAATCTCTGCT--- 171

Db 109 TTCTACATCAGTGAATAACATGATCCAGAGCTTCGACTTTTGGAGTATTCTCTGCTAAG 168  
QY 172 GAAGAAATCTCTATTATTAAGAAACAATCTCTCACTAAATAAGTTTTCGAGCTAAAGAAA 231  
Db 169 GAAGAAAGAAATACAAACAGTTCGAGTGTGGAGCTGAAGCAAAATTTGGAGCGAAGCGCA 228  
QY 232 GATGCTGATATACAAATAAGAAAGACGATTTTACAGAGTAGTCTCCAGGCAATGATTTTCAA 291  
Db 229 ATATCCAACTCTCCCAACCATGATTAATCACTGCTCAAAATTAATCATTTAAATATGAA 288  
QY 292 AATAACTTTAATACAGGATTTTTCAGGAAGTATTGGTTACTCTATGAGCGGACCAAGAAATA 351  
Db 289 AACAAACCGTTTATAGTTTTCAGGAGCTATTGGTTACTCAATGATGATGCTCCAGAAATA 348  
QY 352 GAACCTTGAGCTGCTATACAAATTTAATCCAAATAACCAAGGATTAACAAATGATCTGAT 411  
Db 349 GAGCTTGAGTATCTTATGAACATTTTGTATGTAAGAAATCAAGGTAACAAATTAAGAAT 408  
QY 412 AATGCTGATATACATAAACAATTTTGCATTTATCTCGTAAAGATGCAATG-----GAAGAT 465  
Db 409 GAAGCACATAGATATTGTGCTCTATCCCAATACTCAGCAGCAGACATGATGATGCAAGT 468  
QY 466 CAGCAATATGATGATCTTAAAAATGACGGCATAAAGCTTTTATGCTCATTTGATGTTAATACT 525  
Db 469 AATAATTTTGTCTTTCTAAATAATGAAGATTAATGACATATCATTTATGCTGAACGCA 528  
QY 526 TGCTATGACATTAAGCTGAAGAGTATCTTTTCGATACCATATGCTGATGCTGAGGTATAGA 585  
Db 529 TGCTATGACGTAGTAGGCGAAGGCATACCTTTTCTCTTATATATATGCGCAGGTATCGGT 588  
QY 586 GCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAGGAAATAATA 645  
Db 589 ACTGATTTAGTATCAATGTTGAAGCTACAATCTCAAAATTTCTTACCAGGAAAGTTA 648  
QY 646 GGTATTAGTTACCTATCACACAGAGTCTCTGCAATTTATTTGGTGATACCTACCATGCG 705  
Db 649 GGTTTAAGCTACTCTATAGCCAGAGCTTCTGTGTTTATTTGGTGGCACTTTTCATAAG 708  
QY 706 GTTATTGGTAATAATTTTGAAGATAC---CTGTAATAACTCTCTGATGATTAATAATGAT 762  
Db 709 GTAATAGGAAGCAATTTAGAGATATTCTCTATATAATACCTACTGATCAACACTTGCA 768  
QY 763 GCTCTCTCAACACATCTGCTTTCAGTAACTCTTGGAGTTGGATACCTTTGGCGGAGAAAT 822  
Db 769 GGAAGAGGAATACCTCCCAATAGTAACTGATGATGATGCCACTTTGGAATAGAAATG 828  
QY 823 GGAATGAGGTTC 835  
Db 829 GGAGGAAGGTTTA 841

## RESULT 15

US-09-660-587-43  
; Sequence 43, Application US/09660587  
; Patent No. 6392023

GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.

APPLICANT: Yu, Xue-Jie

TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof

FILE REFERENCE: D6152CIP2

CURRENT APPLICATION NUMBER: US/09/660,587

PRIOR FILING DATE: 2000-09-12

PRIOR APPLICATION NUMBER: 09/261,358

PRIOR FILING DATE: 1999-03-03

NUMBER OF SEQ ID NOS: 46

SEQ ID NO 43

LENGTH: 828

TYPE: DNA

ORGANISM: Ehrlichia canis

FEATURE:

OTHER INFORMATION: nucleic acid sequence of E. canis p28-3

US-09-660-587-43

Query Match 23.2%; Score 195; DB 4; Length 828;

Best Local Similarity 53.9%; Pred. No. 1.4e-41;

Matches 452; Conservative 0; Mismatches 375; Indels 12; Gaps 2;

QY 1 ATGATTTATATAGAAAATTTCTAGTAAGAGCGCGTTAATCTCATTAATGCTCAATCTTACCA 60  
Db 1 ATGAACTGTAAAAAATTTCTTATACAACTACATGTTGGTATCCTAACAAATTTCTTTTACCT 60  
QY 61 TATCAGTCTTTTGGAGATCTCTGTAGTCTCAAGAACTAATGATATAACAAAGAGGCTTTTAC 120  
Db 61 GGCATACTTTCTCCAAACCAATACATGAA-----AACATACTACAGGAACTTTTAC 114  
QY 121 ATTAGTGAAGATGACAACTCAAGTATATCAACATTTAGAAAAATTTCTCTGCTGAAGAACT 180  
Db 115 ATTATTGGAAATATATGATCAAGATATTTCATATTTGGGAACTTTTTCAGCTTAAAGAGAA 174  
QY 181 CCTATTATGAAACAAATTTCTCTCACTAAAAAAGTTTTCGCACTAAAGAAAGATGGTGTAT 240  
Db 175 AAAAACACAACTGGAATTTTGGATTAAAGAAATCATGGACTGGTGGTATCATCTTT 234  
QY 241 ATAACAAAAAAGACGATTTTACAAGAGTAGTCTCCAGGCATTTGATTTTCAAAAATAACTTTA 300  
Db 235 GATAAGAACATGCGAGCTTTTAAATATCCAAATTTATTCATTTAAATATGAAAAATAATCCA 294  
QY 301 ATATCAGATTTTTCAGGAAGTATTGGTTACTCTATGAGACGGACCAAGAAATAGAACTTTGAA 360  
Db 295 TTTTATGATTTGAGGGGTAAATTTGGCTATTTCATATAGGTAGTCCAAAGAAATAGAAATTTGAA 354  
QY 361 GCTCATATCAACAAATTTTAAATCCAAAAAACACCATCAATGATGATGATTAATGGTGAA 420  
Db 355 GTATCATACGAGACATTCGATGTACAAATCCAGAGATAAGTTTAACAATGATGACAT 414  
QY 421 TACTATAAACATTTTGGCAATTTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTAGTA 480  
Db 415 AAGTATTGTGCTTTTATCCAAATGATTCAGTAAACAAATGAAAGTGGTAAATTCGTTTTT 474  
QY 481 CTTAAAAATGACGGATAACTTTTATGTCATGATGGTTAATATCTGCTATGACATTTACA 540  
Db 475 CTCAAAAATGAAGGATTAAGTGACATATCACTCATGTTTAAATGATGTTATGATATATA 534  
QY 541 GCTGAAGAGATCTTTTCGTACCATATGTCATGTCAGGTATAGGAGCAGATCTTATCACT 600  
Db 535 AACAAAGATATGCTTTTTCACCTTACATATGTCAGGCATTTGCTGACTTAATATTC 594  
QY 601 ATTTTAAAGACCTCAATCTAAAAATTTGCTTTACCAAGGAAAAATAGGTATTAGTTACCTT 660  
Db 595 ATGTTTGAAGCTATAAACCATAAAGCTGCTTATCAAGGAAAAATTAGGTTTAAATATATCCA 654  
QY 661 ATCACACAGAGTCTCTGCAATTTTATGGTGGATACCTACCATGCGGTTATTGGTAAATAA 720  
Db 655 ATAAGCCCAAGAGCTAAACATTTCTATGGGTGTGCACTTTCAAAAGTAAACAAACACGAG 714  
QY 721 TTTGAGAGATACCTGTAAATAACTCTCTGATGATTAATGATGCTCTCAAAACCATCT 780  
Db 715 TTTAGAGTTCTGTTTCTAATTAACCTGCTGGAGACT-----CGCTCCAGATTAATCTATT 768  
QY 781 GCTTCAGTAACTCTTTCAGCTTTGGATCTTTTGGCGGAGAAATTTGGAATGAGGTTTCACTTT 839  
Db 769 GCAATAGTAAAGTTGAGTATATGTCATTTTGGTTAGAAATTTGGGTACAGGTCAGTTT 827

## RESULT 16

US-08-733-230-1

; Sequence 1, Application US/08733230

; Patent No. 6025338

GENERAL INFORMATION:

; APPLICANT: Barbet, Anthony F.

; APPLICANT: Ganta, Roman Reddy

; APPLICANT: McGuire, Travis C.

; APPLICANT: Burridge, Michael J.

; APPLICANT: Nyika, Aceme

```

; APPLICANT: Burangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against
; TITLE OF INVENTION: Rickettsial Diseases and Methods of Use
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/733,230
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UP-167
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 864 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..861
; US-08-733-230-1

Query Match 23.2%; Score 195; DB 3; Length 864;
Best Local Similarity 55.4%; Pred. No. 1.4e-41;
Matches 475; Conservative 0; Mismatches 355; Indels 27; Gaps 4;

Qy 1 ATGAATTATAAGAAATTTCTAGTAAGACGCGTTAATCTCATTAATGCTCAATCTTACCA 60
Db 1 ATGAATTCGAAGAAATTTTATCACAAGTACACTAATATCATTAGTGTCTTTTACCT 60

Qy 61 TATCAGTCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGATATACAAGAAAGGCTTCTAC 120
Db 61 GGTGTGTCCTTTTCTGTATGTAATACAGGAAGACAGCAACCCAGC---AGGCAGTGTTTAC 117

Qy 121 ATTAGTCAAAAGTACAATCCAAAGTATATCACACTTTAGAAAATCTCTGCTGGAAGAACT 180
Db 118 ATTAGCGAAATATCATGCCAATCTGCATCATTTTGGTAAATGTCAATCAAGAAGAT 177

Qy 181 CCTATTATGAACAATAATCTCTCACTAAAAA-----GTTTTCGGACTAAAG 228
Db 178 TCAAAAATACTCAACCGTATTTGGTCTAATAAAGATGGGATGGGTTAAACACCA 237

Qy 229 AAAGATGGTATATAACAAAAAGACGATTTTACAGAGTAGTCTCCAGGCAATGATTTT 288
Db 238 TCAGATTCAGCAATCTAATTTCTACAAATTTTACTGAAAAAGACTATTTCTTCAGATAT 297

Qy 289 CAAAATACTTAATATCAGGATTTTTCAGGAAGTATTTGTTACTCTATGCGGACCAAGA 348
Db 298 GAAAAAATCCGTTTTTATAGGTTTCGTCGAGCAATTTGGGTACTCAATGAATGGACCAAGA 357

Qy 349 ATAGAATCTGAAGCTGCTATCACAATTTTAAATCCAAAAAACACCGATAACAATGACT 408
Db 358 ATAGAGTTGGAAGTATCTCATGAACTTTTGATGTAAAAAACCTAGGTGGCACTATAA 417

Qy 409 GATAATGGTGAATACTATAAACAATTT-----TGCATTATCTCGPAAAGATGCAATG 459

```

418 AACACGCACACATGTACTGTGCTTTAGATACAGCAGCACAAAATAGCACTAATGGCGCA 477  
460 GAAGATCAGCAATATGTAGTACTTTAAAAATGACGGCATAACTTTTATGTCATGTATGTT 519  
478 GGATTAACATCATCTGTTATGTTAAAAAAGCAAAATTTAACAAATATATCATTAATGTTA 537  
520 AATATTGCTATGACATTTACAGCTGAAGGAGTATCTTTTCGACCATATGCAATGTCAGGT 579  
538 AATCGTGTATGATATCATCTGTTGATGAATACAGTTCCTCCATATGATGTGCGAGT 597  
580 ATAGGACGATCTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCTTACCACGA 639  
598 ATTGGCAGTACTTAGTGTAGTAAATTAATCTCAAAATCTTAAATTTATCTTATCAAGA 657  
640 AAAATAGGTATTAGTTATCCCTTATCACACCAAGAGTCTCTGCAATTTATTTGGTGGATAC 699  
658 AAGCTAGGCATAGTTACTCAATCAATCTGAAAGCTCTATCTTTATCGGTGGACATTTTC 717  
700 CATGCGTTATTGCTAATAAATTTGAGAAGATACCT---GTAATAACTCTCTGTAGTATTA 756  
718 CATAGAGTTATAGGTAATGAATTTAAAGATATTGCTACCTTAAAAATATTTTACTTCAAA 777  
757 AATCATGCTCTCTCAACACATCTGCTTCAGTAATCTTTCAGTACTCTTTCAGTATCTTTGG 816  
778 ACAGGAATATCTAATCTCTGCGCTTTGCAATCAGCAACACTTGATGTTGTGTCACCTTTG 837  
817 GAAATTTGGAATGAGGTT 833  
838 GAAATTTGAGGAAGTT 854

RESULT 17  
US-08-953-326-1  
; Sequence 1, Application US/08953326  
; Patent No. 6251872  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Ganta, Roman R.  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Nyika, Aceme  
; APPLICANT: Kurangirwa, Fred R.  
; APPLICANT: Mahan, Suman M.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of  
; TITLE OF INVENTION: Animals and Humans  
; FILE REFERENCE: UP-167C1  
; CURRENT APPLICATION NUMBER: US/08/953,326  
; CURRENT FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/953,326  
; EARLIER FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/733,230  
; EARLIER FILING DATE: 1996-10-17  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 864  
; TYPE: DNA  
; ORGANISM: Cowdria ruminantium  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(861)  
; US-08-953-326-1

Query Match 23.2%; Score 195; DB 4; Length 864;  
Best Local Similarity 55.4%; Pred. No. 1.4e-41;  
Matches 475; Conservative 0; Mismatches 355; Indels 27; Gaps 4;

Qy 1 ATGAATTTATAAGAAATTTCTAGTAAGACGCGTTAATCTCATTAATGCTCAATCTTACCA 60  
Db 1 ATGAATTTGCAAGAAATTTTATCACAAGTACACTAATATCATTAGTGTCTTTTACCT 60  
Qy 61 TATCAGTCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGATATACAAGAAAGGCTTCTAC 120  
Db 61 TATCAGTCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGATATACAAGAAAGGCTTCTAC 120

Db 61 GGTGTGCTCTTTCTGTGATGATATACAGGNAGACAGCAACCCAGC---AGGCGAGTGTTCAC 117  
Qy 121 ATTAGTGCAGAGTACAATCAACAGTATATACACTTTAGAAAATTTCTGTGCGAAGAACT 180  
Db 118 ATTAGCGCAAAATACATGCCAAGTGCATCAGATTTTGGTAAATGTCAATCAAGAAGAT 177  
Qy 181 CCTATTATGGAACAAATCTCTCACTAAATAA-----GTTTTCGGACTAAG 228  
Db 178 TCAAAAAATACTCAACCGGTATTTGGTCTAAAAAAGATTTGGATGGCGTTAAAAACACCA 237  
Qy 229 AAGAGTGGTGATATACAAAAAAGAGAGATTTTACAAGAGTAGTCCAGGCAATGATTTT 288  
Db 238 TCAGATCTAGCAATCTAATTTCTCAATTTTCTGAAAAGACTATCTTTTCAGATAT 297  
Qy 289 CAAAAATACTTAATATCAGGATTTTTCAGGAAGTATTTGGTTACTCTATGACGACCAAGA 348  
Db 298 GAAAAACAATCCGTTTTTAGGTTTCGTGGAGCAATTTGGGTACTCAATGAATGGACCAAGA 357  
Qy 349 ATAGAACTTGAAGCTGCATATCAACANTTTATCCAAAAAACACCGATACAGATGACT 408  
Db 358 ATAGAGTTTCGAAGTAGTCTCATGAAACTTTTGATGTAAAAAACCCTAGGTGGCAACTATAA 417  
Qy 409 GATAATGGTGAATACATATAACAATTT-----TGCATTTATCTCGTAAAGATGCAATG 459  
Db 418 AACACGACACATGACTGTGCTTTAGTACAGCAGCAACAATAGCACTAATGGCGA 477  
Qy 460 GAAGATCAGCAATATGTAGTACTTAAAAATGACGGCATAACTTTTATGTCAATTTGATGTT 519  
Db 478 GGATTAACCTACATCTGTTATGTTGTAATAAAGCAAAATTTTAAACAATATATCATTAATGTTA 537  
Qy 520 AATACTTGTATGACATTTACAGCTGAGAGAGTATCTTTTCGTACCATATGTCATGTCAGGT 579  
Db 538 AATGCGTGTATGATATCATGCTTGATGGAATACCAAGTTTCTCCATATGATGTGTCAGGT 597  
Qy 580 ATAGAGCAGAGTCTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGA 639  
Db 598 ATTGCACTGACTTAGTGTAGTCAATTAATGTCACAAATCTCAATATCTTATCAAGGA 657  
Qy 640 AAAATAGTATTAGTTACCTATACACACAGAGTCTCTGCAATTTATTTGGTGGATACACTAC 699  
Db 658 AAGCTAGGCATAAGTTACTCAATCAATCTGAAGCTTCTATCTTTATCGGTGGACATTC 717  
Qy 700 CATGCGCTATTGGTAATAATTTGAGAGATACCT---GTAATACCTCTAGTATTA 756  
Db 718 CATAGAGTTATAGTAAGTAATTTAAAGATATTTGCTACCTTAAAAAATATTTACTTCAAAA 777  
Qy 757 AATGATGCTCTCAAAACACATCTGCTTCAGTAACCTCTTGAGCTTGGATACCTTTGGCGGA 816  
Db 778 ACAGGATATCTAATCCTGGCTTTGCATCAGCAACACTTGTATGTTGTCACTTTGGTATA 837  
Qy 817 GAAATGGGAATGAGGTT 833  
Db 838 GAAATGGAGGAAGGTT 854

## RESULT 18

US-09-660-587-3

Sequence 3, Application US/09660587

Patent No. 6392023

GENERAL INFORMATION:

APPLICANT: Walker, David H.

APPLICANT: McBride, Jere W.

APPLICANT: Yu, Xue-Jie

TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein

TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof

FILE REFERENCE: D6152CIP2

CURRENT APPLICATION NUMBER: US/09/660,587

CURRENT FILING DATE: 2000-09-12

PRIOR APPLICATION NUMBER: 09/261,358

PRIOR FILING DATE: 1999-03-03

NUMBER OF SEQ ID NOS: 46

SEQ ID NO 3

LENGTH: 849

TYPE: DNA

ORGANISM: Ehrlichia canis

FEATURE:

NAME/KEY: mat\_peptide

OTHER INFORMATION: nucleic acid sequence of p28-5

US-09-660-587-3

Query Match 21.6%; Score 181.4; DB 4; Length 849;

Best Local Similarity 53.6%; Pred. No. 4.7e-38;

Matches 458; Conservative 0; Mismatches 376; Indels 21; Gaps 3;

Qy 1 ATGAATTAATAGAAAATTTCTAGTAAGAGCGGTTAATCTCATTAATGTCAATCTTACCA 60  
Db 1 ATGAATTTGAAAAAGTTTTTCAATATAGTCATTTGATATCATTCATATATCTTCTACT 60  
Qy 61 TATCAGTCTTTTTCAGAGATCCTGTAGGTTCAAGAACTAATGATAACAAGAAAGGCTTTTAC 120  
Db 61 AATGCTCATACTCTAACCCAGTATATGGTAACAGTATGTATGTAAT-----TTTTAC 114  
Qy 121 ATTAGTGCAGAAAGTACAATCCAAGTATATCACACTTTAGAAAATTTCTGTGCGAAGAACT 180  
Db 115 ATATCAGGAAAGTACATGCCAAGTGTTCCTCATTTTGGAAATTTTTCAGCTGCAAGAG 174  
Qy 181 CCTATTATGGAAC---AAATTTCTCACTAAAAAAGTTTTTCGGACTAAAGAAAGATGCT 237  
Db 175 AAAAAAAGACACTGTAGTATATGGCTTAAAAAAGAACTGGCGAGGATGCAATATCT 234  
Qy 238 GATATAACAAAAAAGAGCGATTTTCAAGAGTAGTCCAGGCATTTGATTTTCAAAATAC 297  
Db 235 AGTCAAGTCCAGATGATAATTTTACCAATTCGAAATTAATCTCAATCAAGTATGCAAGCAAC 294  
Qy 298 TTAATATCAGAGATTTTCAGGAAGTATTTGGTTACTCTATGAGCAGGACCAAGAAATAGAACTT 357  
Db 295 AAGTTTTTAGGGTTTTGCAGTAGCTATTTGGTTACTCGATAGGCAGTCCAAAGAAATAGAACTT 354  
Qy 358 GAAGTCGATATCAACAATTTAATCCAAAAAACCGATAACAATGATGACTGA----- 410  
Db 355 GAGATGCTTTATGAGCAATTTGATGTGAAAATCCAGGTGATTAATACAAAAACGGTCT 414  
Qy 411 -----TAATGGTGAATACTATAAATTTTCATTTTCATTTTCGTAAGATGCAATGGAAGAT 465  
Db 415 TACAGGTATTGTGCTTTTATCTCATCAAGATGATCGGATGATGACATGACTAGTGCAACT 474  
Qy 466 CAGCAATATGTAGTACTTTAAAAATGACGGCATAACTTTTATGTCATGATGTTAATACT 525  
Db 475 GACAAATTTGATATTTTAAATTAATGAAGGATTAATCAATATCATTTATGACAAACATA 534  
Qy 526 TGCTATGACATTTACAGCTGAAGGAGTATCTTTTCGTACCATATGATGTCAGGATATAGGA 585  
Db 535 TGTATGAAACAGCAAGCAAAAAATATACCTCTCTCTCTTACATATGTGCAAGTATTGGT 594  
Qy 586 GCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAATA 645  
Db 595 ACTGATTTAAATTCACATGTTTGAAGAACTACACATCTCTAAATTTTCTTATCAAGGAAAGCTA 654  
Qy 646 GGTATTAGTTACCTATCACACAGAGAGTCTCTGCAATTTATTTGGTGGATACCTACCATGGC 705  
Db 655 GGGTTGGCTTACTCGTAAGTGCAGAGTCTTTCCGTTTCTTTTGGTATATATTTTCTATAA 714  
Qy 706 GTTATTGGTAAATAATTTGAGAAGATACCTGTAATTAATCTCTGTAGTATTAATATGATGCT 765  
Db 715 ATTATAATATATAGTTTAAAAATGTTCCAGCCATGGTACCTTATTAATCTCAGACGAGATA 774  
Qy 766 CCTCAAAACCATCTGCTTCAGTAACCTTTCAGCGTTGGATATCTTTGGCGGAGAAATTTGGA 825  
Db 775 GTAGGACCAAGTTTGCACACAGTAACATTAATTAATGATGCTACTTTGGATTTAGAACTTGA 834  
Qy 826 ATGAGGTTCACTTTC 840  
Db 835 TGTAGGTTCAACTTTC 849

## RESULT 19

US-09-261-358A-3  
; Sequence 3, Application US/09261358A  
; Patent No. 6403780  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof  
; FILE REFERENCE: D6152CIP  
; CURRENT APPLICATION NUMBER: US/09/261.358A  
; CURRENT FILING DATE: 1999-03-03  
; PRIOR APPLICATION NUMBER: 09/201,458  
; PRIOR FILING DATE: 1998-11-30  
; NUMBER OF SEQ ID NOS: 33  
; SEQ ID NO: 3  
; LENGTH: 849  
; TYPE: DNA  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; OTHER INFORMATION: nucleic acid sequence of ECa28SA2  
US-09-261-358A-3

Query Match 21.6%; Score 181.4; DB 4; Length 849;  
Best Local Similarity 53.6%; Pred. No. 4.7e-38;  
Matches 458; Conservative 0; Mismatches 376; Indels 21; Gaps 3;

Qy 1 ATGAATATAGAAATCTAGTAGAGCGGTATATCTCAATTAATGCTCAATCTTACCA 60  
Db 1 ATGAATTTGAAAAAGTTTTTCCACAATAAGTCATGATATCATATATCTTCTACCT 60  
Qy 61 TATCAGTCTTTTGCAGATCTGTAGTCTCAAGAACTAATGATAACAAGAGGCTTCTAC 120  
Db 61 AATGTCATACTCTAACCAGTATATGGTAACAGTATGTAAGTAAT-----TTTAC 114  
Qy 121 ATTAGTCAAAAGTCAATCCAAAGTATATCACTTTTGAATAATCTCTGCTGAAGAACT 180  
Db 115 ATATCAGAAAGTACATGCCAAGTGTCTCTCATTTTGGAAATTTTTCAGCTGAAGAG 174  
Qy 181 CCTATTATGAAC---AAATCTCTCACTAAAAAGTTTTCGGACTAAAGAGATGGT 237  
Db 175 AAAAAAGACAACTGTATGATATATGGCTTAAAAAGAAACCTGGCAGAGATGCAATATCT 234  
Qy 238 GATATAACAAAAAGACGATTTTACAGAGTAGCTCCAGGCATTTGATTTTCAAAATAAC 297  
Db 235 AGTCNAAGTCCAGATGATAATTTTACCATTTCGAATTTACTCTCAATGATGCAAGCAAC 294  
Qy 298 TTAATATCAGATTTTTCAGGAAGTATTGGTTACTCTATGACCGGACCAAGAAATAGAATT 357  
Db 295 AAGTTTTTAGGGTTTGCAGTAGTATTGGTTACTCGATAGGCAGTCCCAAGAAATAGAGTT 354  
Qy 358 GAAGCTGCATATCAACAATTTAATCCAAAAACACCGATAACATGATGACTGA----- 410  
Db 355 GAGATGCTTATGAAGCAATTTGATGTGAAAAATCCAGGTGATAATTAACAAAAACGGTGCT 414  
Qy 411 -----TAATGTGTAATCTATAAACAATTTTCCATTTATCTCTGTAAGATGCAATGGAAGAT 465  
Db 415 TACAGGTATTGCTTTATCTCATCAAGATGATGCGGATGATGATGATGCTAGTGCAACT 474  
Qy 466 CAGCAATATGATGACTTTAAAAAATGACGGCAATACTTTTATGTCAATGATGTTTAATCT 525  
Db 475 GACAAATTTGATATATTAATTAATGAAGGATTAATCAATATCATTTATGACAAACATA 534  
Qy 526 TGCTATGACATACAGTGAAGGAGTATCTTTTGTACCATATGATGATGCAAGGTATAGGA 585  
Db 535 TGTATGAAACAGCAAGCAAAAAATATACCTCTCTCTTACATATGTCAGGATTTGGT 594  
Qy 586 GCAGATCTTATCACTATTTTTAAAGACCTCAATCTAAAAATTTGCTTCAAGAGGAAATA 645  
Db 595 ACTGATTAATTCATGTTTTGAAACTACATCATCTTAAATTTCTTATCAAGGAAGCTA 654

Qy 646 GGTAATTAGTACCCTATCATCACACAGAAAGTCTCTGCATTTATTTGGTGGATACCTACATGGC 705  
Db 655 GGGTTGGCTACTTCGTAAGTGCAGAGTCTTCGGTTTCTTTTGGTATATATTTTCATAAA 714  
Qy 706 GTTATTGGTAATAATTTGAGAGATACCTGTATAACTCTCTGTAGTATTAAATGATGCT 765  
Db 715 ATTATAAATAAATAGTTTAAAAATGTTCCAGCCATGGTACCTATTAACTCAGACGAGATA 774  
Qy 766 CCTCAAAACCATCTGCTTCCAGTAACTCTTGACCTTGGACTTTTGGCGGAGAAATGGA 825  
Db 775 GTAGGACCACAGTTTGCACACAGTAACATTAATGTATGCTACTTTGGATTAGAACTTGA 834  
Qy 826 ATGAGGTTCACTTC 840  
Db 835 TGTAGGTTCACTTC 849

## RESULT 20

US-08-953-326-12  
; Sequence 12, Application US/08953326  
; Patent No. 6251872  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Ganta, Roman R.  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Burrige, Michael J.  
; APPLICANT: Nyika, Aceme  
; APPLICANT: Ruxangirwa, Fred R.  
; APPLICANT: Mahan, Sunan M.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of  
; TITLE OF INVENTION: Animals and Humans  
; FILE REFERENCE: UP-167C1  
; CURRENT APPLICATION NUMBER: US/08/953.326  
; CURRENT FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/953.326  
; EARLIER FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/733.230  
; EARLIER FILING DATE: 1996-10-17  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 12  
; LENGTH: 864  
; TYPE: DNA  
; ORGANISM: Ehrlichia canis  
US-08-953-326-12

Query Match 21.0%; Score 176.8; DB 4; Length 864;  
Best Local Similarity 55.0%; Pred. No. 7.3e-37;  
Matches 480; Conservative 0; Mismatches 347; Indels 45; Gaps 5;

Qy 1 ATGAATATAGAAAAATTTCTAGTAAGAGCGGTTAATCTCAATTAATGCTCAATCTTACCA 60  
Db 3 ATGAATATAGAAAAATTTTACAGTAACCTGTCATTAGTATTATTTACTTCTTTACACAT 62  
Qy 61 TATCAGCTTTTGCAGATCTCTGTAGGTTCAAGAACTAATGATACAAAGAGGCTTCTAC 120  
Db 63 TTTATACCTTTTATAGTCCAGCAGCGTCCAGTACAATTCACAA-----CTTCTAC 113  
Qy 121 ATTAGTCAAGTACATCAATCCAGTATATCACTTTTAGAAAAATTTCTGCT---GAAGAA 177  
Db 114 ATTAGTGGAAAAATATATGCAACAGCGTCACTTTTGGAAATTTTTCAGCTTAAAGAA 173  
Qy 178 ACTCTATTAAATGAACAAATTTCTCACTAAAAAGTTTTCGACCTAAAGAAAGATGGT 237  
Db 174 CAAGTTTACTAAGGTATTAGTTGGTTAGATCAACGATTTATCAATATATTATTAAC 233  
Qy 238 GATATAACAAAAAGACGATTTTACAAGAGTAGTCCAGGCATTTGATTTTCAAAAAATAC 297  
Db 234 AATAATCATACAGCAAGAGCTTTAAGGTTCAAAATTTATCTTTAAATACAAAAATAAC 293  
Qy 298 TTAATATCAGATTTTTCAGGAAGTATTGGTTACTCTATGACCGGACCAAGAAATAGAATT 357  
Db 294 CCATTTCTAGGATTTGAGGAGCTATTGGTTATTTCATAGGCAATTCAGAAATAGAACTA 353

358	Qy	GAAGCTGCATATCAACAATTTTAATCCAAAAACACCGATAACAANTGATACTGATAATGGT	4117
354	Db	GAAAGTATCACATGAAATATTTTGATACTAAAAACCCAGGAACAATTTATTTAAATGACTCT	4113
418	Qy	GAATACTATAAACATT-----TTGCATTATCTCGTNA	449
414	Db	CACAAATATTGCGCTTTATCTCATGGAAGTCACATATGCAGTGATGGAATAGCGGAGAT	473
450	Qy	AGATGCAATGGAAGA--TCAGCAATATGTAGTACTTTAAAAATGACGGCATACTTTTATG	507
474	Db	TGGTACACTGCAAAAACATGATAAGTTGTACTCTTCGAAAAATGAAGTTTACTTGACGTC	533
508	Qy	TCATTGATGGTTAATPACTTGGCTATCACATTAACAGCTGAAGGAGTACTTTTCGTACCATAT	567
534	Db	TCATTTATGTTAAACGCATGTTATGACATAACAACCTGAAAAAATGCCITTTTTCACCTTAT	593
568	Qy	GCATGTGCAGGTATAGGAGCAGATCTTATCATCTTTTAAAAAGACCTCAATCTAAAAATT	627
594	Db	ATATGTGCAGGTATTTGGTACTGATCTCATATCTATGTTTGAGACAACAACAAAAATA	653
628	Qy	GCATTACCAAGAAAAATAGGTATTAGTTTACCCTATCACACGAGAGTCTTCGATTTATT	687
654	Db	TCTTTATCAAGAAAATTTAGGTTTAAACATATACATAAACTCAAGAGTTTCTGTTTTGCA	713
688	Qy	GGTGATACTACCATGGCGTTATTTCGTAATAAATTTGAGAAGATACCTGTAATACTCT	747
714	Db	GGTGGCACTTTTCAAGGTATAGGTATAGTAATTTAAAGGTATTCCTACTCTATTACCT	773
748	Qy	GTAGTATTTAAATGATGCTCCTCAAAACCATCTGCTTTCAGTTAACTCTTGACGTTGGATAC	807
774	Db	GATGGATCAACATTTAAAGTACAA--CAGTCTGCAACAGTAACATTAGATGTGTGCCAT	830
808	Qy	TTTGGCGGAGAAATTTGGAATGAGGTTTCACCTT	839
831	Db	TTCCGGGTTAGAGATTTGGAAATGAGATTTTCTT	862

## RESULT 21

```

US-09-660-587-45
; Sequence 45, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152C3P2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 45
; LENGTH: 813
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: nucleic acid sequence of E. canis p28-9
US-09-660-587-45

```

	Query Match	17.5%	Score 147.2;	DB 4;	Length 813;
	Best Local Similarity	53.2%	Pred. No. 3.2e-29;		
	Matches	445;	Conservative	0; Mismatches 358;	Indels 33; Gaps 5
<hr/>					
Qy	1	ATGAATTATAAGAAAATTC	TAGTAAAGAAGCGGTTAATCTCATTAAATGTCAATCTTACCA	60	
Db	1	ATGAATTACAAAGATTGGT	TGGTAGGTCTTACGCTGAGTCATATGTTTTTCTTTCTTAICT	60	
<hr/>					
Qy	61	TATCAGTCTTTTGAGACTCC	TGTAGGTTCAAGAACTAATGATAACAAGAAGAGGCTTTTAC	120	
Db	61	GATGGTGCTTTTCTCTGA	-----TGCAAATTTTCTGAAGGAGGAGAGGAGCTTTAT	111	

Qy	121	ATTAGTGAAGAGTACAATCCAAAGTATATACACATTTTGAAGAAATTTCTCTGCTGAAGAAACT	180
Db	112	ATAGGTAGTCAGTATAAAGTTGGTATTCCCAATTTTGTAGTAATTTTTCAGCTGAAGAAACA	171
Qy	181	CCTATTAAATGGAAACAATTCCTCACATAAAAAAGTTTTTCGGACATAAAGAAAGATGGTGAT	240
Db	172	ATTCCTGGTATTACAAAAGAGATTTTTCGGTTAGGTCTTG-----ATAAGTCTGAG	222
Qy	241	ATAACAAAAAAGACGATTTTACAAGAGTAGCTCCAGCAATGATTTTTCACAAATTAACATTA	300
Db	223	ATAAATACTCACAGCAATTTTACACGATCATATGACCCCTACTTAT-----GCAAGCAGT	276
Qy	301	ATATCAGATTTTCAGGAAGTATTTGGTTACTCTATGGACGGACCAAGAATAAGACTTCGAA	360
Db	277	TTTCGAGGGTTTAGTGGTATCAITGGATATTATGTTAATGACTTTTAGGGTAGAATTTGAA	336
Qy	361	GCTGCATATCAACAATTTAAATCCAAAAAACCCGATAAACAATGATCTATGATAATGGTGAA	420
Db	337	GGTTCCTATGAGAAATTTTGAACCTGAAAGACAATGGTACC-----CTGAGAATAGCCAA	390
Qy	421	TACTATAAACATTTTTCGATTTATCTGTAAAGATGCAATGGGAAGATCAGCAATACTAGTA	480
Db	391	AGCTACAAATTTTTCGTTCTGCGAAATGCTCAAAATAGTGATAATAAGTTTATAGTA	450
Qy	481	CTTAAAAATGACGGCATAACTTTTATGTCAATGTATGGTTAATCTTGCTATGACATTACA	540
Db	451	CTAGAGAATAACGGGGTTGTGACAAGTCTCTTAATGTAAATGTTTGTATGATATTGCT	510
Qy	541	GCTGAAGAGTATCTTTCGTACCAATATGCATGTGCAGGTATAGGAGCAGATCTTATCACT	600
Db	511	AGTGGTAGTATTCCTTTTAGCACCTTATATGTGTCTGGTGTGGTGGCAGATATATAAAG	570
Qy	601	ATTTTTAAAGACCTCAATCTAAAAATTTGCTTTACCAAGGAAAAATAGGTATTAGTTACCC	660
Db	571	TTTTTAGGTATACATTCGCCCTAAGTTTCTTATCAAGTTAAGTTTGGTGTCACTACCC	630
Qy	661	ATCACACAGAGTCTCTGCATTTTATGGTGGATACATACCATGGCGGTTATTGGTAAATAAA	720
Db	631	CTAAATGTTAATACTATCTGTTGTTGGTGGGGGTTATTACCATAAGGTTGTAGGTGATAGG	690
Qy	721	TTTCAGAGATACCTGTA--ATAACTCTCTGATTAAATGATGCTCCCTCAAAACCACA	777
Db	691	CATGAGAGATGAGAAATAGCTTACCATCCTACTGCATTTATCTGACGTTCTCTAGAACTACT	750
Qy	778	TCTGCTTCAGTAACTCTTCAGCTTTGGATCTTTTGGCGGAGAAATTTGGAATGAGTTT	833
Db	751	TCAGCTTCGTCTACTTTTAAATCACTGATTTATTTGGTTGGGAGATGGGATTTAGATT	806

RESULT 22

RES001. 22  
US-09-660-587-39  
; Sequence 39, Application US/09660587  
; Patent NO. 6392023  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof  
; FILE REFERENCE: D6152C1P2  
; CURRENT APPLICATION NUMBER: US/09/660,587  
; CURRENT FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 39  
; LENGTH: 879  
; TYPE: DNA  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: nucleic acid sequence of E. canis p28-1  
US-09-660-587-39

Query Match 13.2%; Score 110.8; DB 4; Length 879;  
Best Local Similarity 56.6%; Pred. No. 8.5e-20;  
Matches 205; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 478 GTACTTAAATGAGCGCATACCTTTTATGTCATTGATGTTTAACTTCTGCTATGACATT 537  
DB 517 GTAATGAGAAATGATGGATTATCTATAATATCTGTTATAGTAAATGTTTGTCTGATTC 576

QY 538 ACAGCTGAAGGAGTATCTTTCGTACCAATATGATGTCAGGTATAGGAGCAGATCTTATC 597  
DB 577 TCTTTGAACAATTTGCTCAATATGCTTTACATAATGTCGAGGAGCGAGGTAGTCTATA 636

QY 598 ACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAATAGGTATTAGTTTAC 657  
DB 637 GAATTTCTTCGATGATTACACATTAAGTTTGCATATCAAGCAAGCTAGGTATTGCTTAT 696

QY 658 CCTATCACACAGAGTCTCTGCATTTTATGCTGGATACCTACCATGGGTTTATGGTAAAT 717  
DB 697 TCTCTACATCTAACATTAGTCTCTTTCGTAGTTTATATTAACATAAAGTAAATGGGCAAT 756

QY 718 AAATTTGAGAAGATACCTGTAAATCTCTGTAGTATTAAATGATGCTCTCAAAACACA 777  
DB 757 CAATTTAAATTTAAATGTCACATGTTCTGAATTTGCAAGTATACCTAAATTTACA 816

QY 778 TCTGTTTCAATCTTTGAGCTTTGATGATCTTTGGCGGAGAAATTTGGAATGAGGTTCAAC 837  
DB 817 TCCGAGTTGCTACACTTAATATTGTTTATTTTGGAGTGAAATTTGTCGCAAGATTGACA 876

QY 838 TT 839  
DB 877 TT 878

## RESULT 23

US-08-953-326-13  
; Sequence 13, Application US/08953326  
; Patent No. 6251872  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Ganta, Roman Reddy  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Nyika, Aceme  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Mahan, Suman M.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of  
; FILE REFERENCE: UP-167C1  
; CURRENT APPLICATION NUMBER: US/08/953,326  
; CURRENT FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/953,326  
; EARLIER FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/733,230  
; EARLIER FILING DATE: 1996-10-17  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 399  
; TYPE: DNA  
; ORGANISM: Ehrlichia canis  
US-08-953-326-13

Query Match 10.9%; Score 91.4; DB 4; Length 399;  
Best Local Similarity 55.3%; Pred. No. 6.8e-15;  
Matches 223; Conservative 0; Mismatches 171; Indels 9; Gaps 2;

QY 1 ATGAATTATAAGAAAATTTCTAGTAAGAGCGCGTTAATCTCAATTAATGTCATCTTACCA 60  
DB 3 ATGAATTGTAAGAAAATTTTCAATAAGTGCATTGATATCATCATATCTTCTACCT 62  
QY 61 TATCAGTCTTTGAGATCCCTGTAGTTCAGAACTAATGATAACAAGAGGCTTCTAC 120

DB 63 AATGCTCATACTCTTAACCCAGTATATGTTAAACAGTATGTTGTAAT-----TTTTAC 116  
QY 121 ATTAGTCAAAAGTACAAATCCAAAGTATATCAACATTTTAGAAAATTTCTCTGCTGAAGAACT 180  
DB 117 ATATCAGAAAAGTACATGCCAAGTGTCTCTCATTTTGGAAATTTTTCAGCTGAAGAG 176  
QY 181 CTTATTAATGGAAC---AAATTTCTCTCACTTAAAGTGTTCGGACTTAAAGAAAGATGGT 237  
DB 177 AAAAAAAGACAACCTAGTATATATGCTTAAAGAAAACCTGGCGAGGAGATGCAATATCT 236  
QY 238 GATATACAAAAAAGACGATTTTACAGAGTAGTCTCCAGGCATTTGATTTTCAAAATAAC 297  
DB 237 AGTCAAAAGTCAGATGATAATTTTACCATTTCGAAATTAATCTCAATTCAGATGCAAGCAAC 296  
QY 298 TTAATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGACGCGACCAAGAAATAGAACTT 357  
DB 297 AAGTTTTTAGGGTTTGCAGTAGCTATTGGTTACTTCGATAGCCAGTCCCAAGATAGAGTT 356  
QY 358 GAAGCTCATATCAACAATTTAATCCAAAAAACACCGATTAACA 400  
DB 357 GAGATGCTTATGAAGCATTTGATGTAATAAAATCAAGGTAACA 399

## RESULT 24

US-08-733-230-5  
; Sequence 5, Application US/08733230  
; Patent No. 6025338  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Ganta, Roman Reddy  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Nyika, Aceme  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Mahan, Suman M.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against  
; TITLE OF INVENTION: Rickettsial Diseases and Methods of Use  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/733,230  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Whitlock, Ted W.  
; REGISTRATION NUMBER: 36,965  
; REFERENCE/DOCKET NUMBER: UP-167  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 352-375-8100  
; TELEFAX: 352-372-5800  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 849 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1...846  
US-08-733-230-5

Query Match 10.8%; Score 91; DB 3; Length 849;  
Best Local Similarity 50.5%; Pred. No. 1.1e-14;  
Matches 251; Conservative 0; Mismatches 240; Indels 6; Gaps 1;

QY 281 TTGATTTTCAAAATAAATTAATATCAGGATTTTTCAGGAAGTATTTGGTTACTCTATGGACG 340  
DB 299 TTGCTTCTCTAAATAAATTAATATCAGGATTTTTCAGGAAGTATTTGGTTACTCTATGGACG 358

QY 341 GACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAACACCGATACA 400  
DB 359 GAGCAGAGTGGAAATTTGAAGCGAGCTACAGAAGGTTTGCTACTTTGGCGGACGGCGAGT 418

QY 401 ATGATCTGATAATGGTGAATTAATCAATTTTGAATTTTGCATTTCTCGTAAGATGCAATGG 460  
DB 419 ACGCAAAAGTGGTGGGAAT-----CTCTGGCAGCTATTACCGCGACGCTAACATTA 472

QY 461 AAGATCAGCAATATCTAGTACTTTAAATATGAGCGCAATACTTTTATGTCATTTGATGGTTA 520  
DB 473 CTGAGACCAATTTACTTCTGTAGTCAAAATTTGATGAATCAACAACACCTCAGTCAATGTTAA 532

QY 521 ATACTTGTATGACATTTACAGCTGAAGAGTATCTTTTGTACCATATGATGTCAGGTA 580  
DB 533 ATGGCTGTATGACGTGTGCACAGATTTTACCTGTGTCCCGGTATGTATGTGCGGGA 592

QY 581 TAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTTGTCTTACCAAGAA 640  
DB 593 TAGGCGCAGCTTTGTGACATCTCTAAGCAAGTAAACCAAGCTGGCTTACAGGGCA 652

QY 641 AAATAGGTATTAGTTACCTATCACACAGAGTCTCTGCAATTTATTTGGTGGATACTACC 700  
DB 653 AGTTGGGATTTAGCTACCAAGTTTACTCCGGAATATCTCTGGTGGCAGGTGGTTCTACC 712

QY 701 ATGGCGTTATTTGGTAAATAATTTGAGAGATACCTGTATATACTCTGTAGTATTAAATG 760  
DB 713 ACGGGCTATTGATGAGTCTTACAAGGACATTTCCCGGCACACACAGTGTAAAGTTCTCTG 772

QY 761 ATGCTCTCTCAACACCACA 777  
DB 773 GAGAAGCAAAAGCCTCA 789

RESULT 25  
US-08-953-326-5  
; Sequence 5, Application US/08953326  
; Patent No. 6251872  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Ganta, Roman R.  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Nyika, Aceme  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Mahan, Suman M.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of  
; TITLE OF INVENTION: Animals and Humans  
; FILE REFERENCE: UF-167C1  
; CURRENT APPLICATION NUMBER: US/08/953,326  
; CURRENT FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/953,326  
; EARLIER FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/733,230  
; EARLIER FILING DATE: 1996-10-17  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 849  
; TYPE: DNA  
; ORGANISM: Anaplasma marginale  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(846)  
US-08-953-326-5

Query Match 10.8%; Score 91; DB 4; Length 849;  
Best Local Similarity 50.5%; Pred. No. 1.1e-14;  
Matches 251; Conservative 0; Mismatches 240; Indels 6; Gaps 1;

QY 281 TTGATTTTCAAAATAAATTAATATCAGGATTTTTCAGGAAGTATTTGGTTACTCTATGGACG 340  
DB 299 TTGCTTCTCTAAATAAATTAATATCAGGATTTTTCAGGAAGTATTTGGTTACTCTATGGACG 358

QY 341 GACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAACACCGATACA 400  
DB 359 GAGCAGAGTGGAAATTTGAAGCGAGCTACAGAAGGTTTGCTACTTTGGCGGACGGCGAGT 418

QY 401 ATGATCTGATAATGGTGAATTAATCAATTTTGAATTTTGCATTTCTCGTAAGATGCAATGG 460  
DB 419 ACGCAAAAGTGGTGGGAAT-----CTCTGGCAGCTATTACCGCGACGCTAACATTA 472

QY 461 AAGATCAGCAATATCTAGTACTTTAAATATGAGCGCAATACTTTTATGTCATTTGATGGTTA 520  
DB 473 CTGAGACCAATTTACTTCTGTAGTCAAAATTTGATGAATCAACAACACCTCAGTCAATGTTAA 532

QY 521 ATACTTGTATGACATTTACAGCTGAAGAGTATCTTTTGTACCATATGATGTCAGGTA 580  
DB 533 ATGGCTGTATGACGTGTGCACAGATTTTACCTGTGTCCCGGTATGTATGTGCGGGA 592

QY 581 TAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTTGTCTTACCAAGAA 640  
DB 593 TAGGCGCAGCTTTGTGACATCTCTAAGCAAGTAAACCAAGCTGGCTTACAGGGCA 652

QY 641 AAATAGGTATTAGTTACCTATCACACAGAGTCTCTGCAATTTATTTGGTGGATACTACC 700  
DB 653 AGTTGGGATTTAGCTACCAAGTTTACTCCGGAATATCTCTGGTGGCAGGTGGTTCTACC 712

QY 701 ATGGCGTTATTTGGTAAATAATTTGAGAGATACCTGTATATACTCTGTAGTATTAAATG 760  
DB 713 ACGGGCTATTGATGAGTCTTACAAGGACATTTCCCGGCACACACAGTGTAAAGTTCTCTG 772

QY 761 ATGCTCTCTCAACACCACA 777  
DB 773 GAGAAGCAAAAGCCTCA 789

RESULT 26  
US-09-288-339-1  
; Sequence 1, Application US/09288339  
; Patent No. 6436399  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Zhi, Ning  
; APPLICANT: Ohashi, No. 6436399io  
; TITLE OF INVENTION: Nucleic Acid Encoding the Major Outer Membrane Protein of  
; TITLE OF INVENTION: the Causative Agent of Human Granulocytic Ehrlichiosis  
; TITLE OF INVENTION: and Peptides Encoded Thereby  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/09/288,339  
; CURRENT FILING DATE: 1999-04-08  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 1242  
; TYPE: DNA  
; ORGANISM: P44  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1242)  
US-09-288-339-1

Query Match 8.8%; Score 74.2; DB 4; Length 1242;  
Best Local Similarity 58.4%; Pred. No. 2.7e-10;  
Matches 149; Conservative 0; Mismatches 103; Indels 3; Gaps 1;

QY 474 TGTAGTACTTAAATAATGACGGCAATACTTTTATGTCATTTAGTGGTTAATCTGCTATGA 533  
DB 960 TGAGGTTGTAGAAATTTAGGCGGGTTCTTCTACTTCTGTAATGGTCAATGCTTGTATGA 1019





```
; FILE REFERENCE: 210121.439C4
; CURRENT APPLICATION NUMBER: US/09/295,028
; CURRENT FILING DATE: 1999-04-20
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 40
; LENGTH: 1919
; TYPE: DNA
; ORGANISM: Ehrlichia sp.
US-09-295-028-40

Query Match      8.8%; Score 74.2; DB 4; Length 1919;
Best Local Similarity 58.4%; Pred. No. 3.2e-10;
Matches 149; Conservative 0; Mismatches 103; Indels 3; Gaps 1;

QY 474 TGTAGTACTTAAATAATGACGGCATAAATCTTTTATGTCATGATGTTAATCTTCTATGA 533
    |||||
Db 383 TGAAGTTGTTGAGATCAGGGCGGTTCTTCTACTTCTGATGTTAATCTTCTATGA 442
    |||||
QY 534 CATTACAGCTCAAGAGAGTATCTTTTCGTACCATATGTCAGTATGAGGAGAGATCT 593
    |||||
Db 443 TCTTCTTAGTGAAGGTTTAGCGGTTGTTCTTACGCTTGTGTCGCTTCGGAGGTAACCT 502
    |||||
QY 594 TATCACTATTTTTAAAGACC---TCAATCTAAATTTGCTTACCAAGGAAAAATAGGTAT 650
    |||||
Db 503 CGTGGCGGTTGTTGATGGCATATACCTCTAAGCTTGTCTATAGATTAAAGGCTGGCTT 562
    |||||
QY 651 TAGTTACCTATACACAGAGTCTCTGCAATTTATTTGTTGGATACCTACCATGGCTTAT 710
    |||||
Db 563 GAGTTATCAGCTCTCTCTGAAATCTCTGCTTTTCTGCTGGGGGTTTCTACCATCGTGTGT 622
    |||||
QY 711 TGGTAATAAATTGA 725
    |||||
Db 623 GGGAGATGGTGTGTTA 637

RESULT 31
US-09-288-339-5
; Sequence 5, Application US/09288339
; Patent No. 6436399
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Zhi, Ning
; APPLICANT: Ohashi, No. 6436399io
; TITLE OF INVENTION: Nucleic Acid Encoding the Major Outer Membrane Protein of
; TITLE OF INVENTION: The Causative Agent of Human Granulocytic Ehrlichiosis
; TITLE OF INVENTION: and Peptides Encoded Thereby
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/288,339
; CURRENT FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patent in ver. 2.0
; SEQ ID NO 5
; LENGTH: 1176
; TYPE: DNA
; ORGANISM: P44-12
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1176)
US-09-288-339-5

Query Match      8.6%; Score 72.6; DB 4; Length 1176;
Best Local Similarity 58.0%; Pred. No. 7e-10;
Matches 148; Conservative 0; Mismatches 104; Indels 3; Gaps 1;

QY 474 TGTAGTACTTAAATAATGACGGCATAAATCTTTTATGTCATGATGTTAATCTTCTATGA 533
    |||||
Db 894 TGAAGTTGTTGAGATCAGGGCGGTTCTTCTACTTCCGTATGTTCAATGCTTGTATGA 953
    |||||
QY 534 CATTACAGCTCAAGAGAGTATCTTTTCGTACCATATGTCAGTATGAGGAGAGATCT 593
    |||||
Db 954 TCTTCTTAGTGAAGGTTTAGGTGTTGTTCTTATGTTGTTGTTGTTGTTGTTGTTGTT 1013
    |||||
QY 594 TATCACTATTTTTAAAGACC---TCAATCTAAATTTGCTTACCAAGGAAAAATAGGTAT 650
    |||||
Db 1014 CGTGGCGGTTGTTGATGGCCATATACCTCTTACGCTTGTCTTATAGATTAAAGGCTGGTT 1073
    |||||
QY 651 TAGTTACCTATACACAGAGTCTCTGCAATTTATTTGTTGGATACCTACCATGGCTTAT 710
    |||||
```

```
; FILE REFERENCE: 210121.439C4
; CURRENT APPLICATION NUMBER: US/09/295,028
; CURRENT FILING DATE: 1999-04-20
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 40
; LENGTH: 1919
; TYPE: DNA
; ORGANISM: Ehrlichia sp.
US-09-295-028-40

Query Match      8.8%; Score 74.2; DB 4; Length 1919;
Best Local Similarity 58.4%; Pred. No. 3.2e-10;
Matches 149; Conservative 0; Mismatches 103; Indels 3; Gaps 1;

QY 474 TGTAGTACTTAAATAATGACGGCATAAATCTTTTATGTCATGATGTTAATCTTCTATGA 533
    |||||
Db 383 TGAAGTTGTTGAGATCAGGGCGGTTCTTCTACTTCTGATGTTAATCTTCTATGA 442
    |||||
QY 534 CATTACAGCTCAAGAGAGTATCTTTTCGTACCATATGTCAGTATGAGGAGAGATCT 593
    |||||
Db 443 TCTTCTTAGTGAAGGTTTAGCGGTTGTTCTTACGCTTGTGTCGCTTCGGAGGTAACCT 502
    |||||
QY 594 TATCACTATTTTTAAAGACC---TCAATCTAAATTTGCTTACCAAGGAAAAATAGGTAT 650
    |||||
Db 503 CGTGGCGGTTGTTGATGGCATATACCTCTAAGCTTGTCTATAGATTAAAGGCTGGCTT 562
    |||||
QY 651 TAGTTACCTATACACAGAGTCTCTGCAATTTATTTGTTGGATACCTACCATGGCTTAT 710
    |||||
Db 563 GAGTTATCAGCTCTCTCTGAAATCTCTGCTTTTCTGCTGGGGGTTTCTACCATCGTGTGT 622
    |||||
QY 711 TGGTAATAAATTGA 725
    |||||
Db 623 GGGAGATGGTGTGTTA 637

RESULT 30
US-09-106-582-40
; Sequence 40, Application US/09106582
; Patent No. 6306402
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; FILING DATE: 29-JUN-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.439C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1919 base pairs
; TYPE: nucleic acid
THERAPY
```

Db 1074 GAGTTATCAGCTCTCTCTGTAATCTCGCTTTTGGCGGTGGATCTTACCATCGGTTGT 1133

Qy 711 TGGTAATAAATTGA 725

Db 1134 GGGAGATGGCGTTTA 1148

RESULT 32

US-09-066-046-26

; Sequence 26, Application US/09066046A

; Patent No. 6204252

; GENERAL INFORMATION:

; APPLICANT: MURPHY, Cheryl

; STOREY, James

; BELTZ, Gerald A.

; COUGHLIN, Richard T.

; TITLE OF INVENTION: CHARACTERIZATION OF GRANULOCYTIC

; EHRlichia AND METHODS OF USE

; NUMBER OF SEQUENCES: 65

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: HALE AND DORR LLP

; STREET: 60 State Street

; CITY: Boston

; STATE: MA

; COUNTRY: United States

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/066,046A

; FILING DATE: 24-Apr-1998

; CLASSIFICATION: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Superko, Colleen

; REGISTRATION NUMBER: 39,850

; REFERENCE/DOCKET NUMBER: 106.941.155

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 526-6000

; TELEFAX: (617) 526-5000

; INFORMATION FOR SEQ ID NO: 26:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1840 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; SEQUENCE DESCRIPTION: SEQ ID NO: 26:

US-09-066-046-26

Query Match 8.6%; Score 72.6; DB 4; Length 1840;

Best Local Similarity 58.0%; Pred. No. 8.1e-10;

Matches 148; Conservative 0; Mismatches 104; Indels 3; Gaps 1;

Qy 474 TGTAGTACTTAAATGACGCGCATATCTTTTATGTCATTGATGGTTAATCTTCTGATGA 533

Db 1056 TGAAGTTGTTGAGATCAGGCGGTTTCTTCTACTTCGGTAATGGTCAATGCTTGTATGA 1115

Qy 534 CATTACAGCTGAAGGAGTATCTTTTCGTACCATATGTCATGTCAGGTATAGGAGCAGATCT 593

Db 1116 TCTTCTTAGTCAAGTTTAGGTGTTGTTCTTATGCTTGTGTTGTTGTTGTTGTTGTTGTT 1175

Qy 594 TATCAGCTATTTTAAAGACC---TCAATCTAAATTTGCTTACCAAGGAAAAATAGGTAT 650

Db 1176 CGTGGCGGTGTTGATGGCCATATCACTCTTAAGCTTGTCTTATAGATTAAAGGCGTGGTT 1235

Qy 651 TAGTTACCTATCACACAGAGTCTCTGCAATTTATGTTGATACCTACCATGGGTTAT 710

Db 1236 GAGTTATCAGCTCTCTCTGTAATCTCGCTTTTGGCGGTGGATTTCTACCATCGGTTGT 1295

Qy 711 TGGTAATAAATTGA 725

Db 1296 GGGAGATGGCGTTTA 1310

RESULT 33

US-09-066-047-12/c

; Sequence 12, Application US/09066047A

; Patent No. 6306394

; GENERAL INFORMATION:

; APPLICANT: MURPHY, Cheryl

; STOREY, James

; BELTZ, Gerald A.

; COUGHLIN, Richard T.

; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND METHODS OF

; USE OF GRANULOCYTIC EHRlichia

; NUMBER OF SEQUENCES: 41

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: HALE AND DORR LLP

; STREET: 60 State Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: United States

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/066,047A

; FILING DATE: 24-Apr-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/044,869

; FILING DATE: 25-APR-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Superko, Colleen

; REGISTRATION NUMBER: 39,850

; REFERENCE/DOCKET NUMBER: 106.941.156

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 526-6000

; TELEFAX: (617) 526-5000

; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1843 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; SEQUENCE DESCRIPTION: SEQ ID NO: 12:

US-09-066-047-12

Query Match 8.6%; Score 72.6; DB 4; Length 1843;

Best Local Similarity 58.0%; Pred. No. 8.1e-10;

Matches 148; Conservative 0; Mismatches 104; Indels 3; Gaps 1;

Qy 474 TGTAGTACTTAAATGACGCGCATATCTTTTATGTCATTGATGGTTAATCTTCTGATGA 533

Db 788 TGAAGTTGTTGAGATCAGGCGGTTTCTTCTACTTCGGTAATGGTCAATGCTTGTATGA 729

Qy 534 CATTACAGCTGAAGGAGTATCTTTTCGTACCATATGTCATGTCAGGTATAGGAGCAGATCT 593

Db 728 TCTTCTTAGTCAAGTTTAGGTGTTGTTCTTATGCTTGTGTTGTTGTTGTTGTTGTTGTT 669

Qy 594 TATCAGCTATTTTAAAGACC---TCAATCTAAATTTGCTTACCAAGGAAAAATAGGTAT 650

Db 668 CGTGGCGGTGTTGATGGCCATATCACTCTTAAGCTTGTCTTATAGATTAAAGGCGTGGTT 609

Qy 651 TAGTTACCTATCACACAGAGTCTCTGCAATTTATGTTGATACCTACCATGGCGTTAT 710

Db 608 GAGTATCAGCTCTCTCTGTAATCTCCGCTTTTGGGGGAGTCTTATCATCGCTGT 549  
Qy 711 TGGTAATAAATTGA 725  
Db 548 GGGAGATGGCGTTTA 534

RESULT 34  
US-09-066-046-28  
; Sequence 28, Application US/09066046A  
; Patent No. 6204252  
; GENERAL INFORMATION:  
; APPLICANT: MURPHY, Cheryl  
; STOREY, James  
; BELTZ, Gerald A.  
; COUGHLIN, Richard T.  
; TITLE OF INVENTION: CHARACTERIZATION OF GRANULOCYTIC  
; EHRlichIA AND METHODS OF USE  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HALE AND DORR LLP  
; STREET: 60 State Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: United States  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/066,046A  
; FILING DATE: 24-Apr-1998  
; CLASSIFICATION: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Superko, Colleen  
; REGISTRATION NUMBER: 39,850  
; REFERENCE/DOCKET NUMBER: 106.941.155  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 526-6000  
; TELEFAX: (617) 526-6000  
; INFORMATION FOR SEQ ID NO: 28:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3435 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; SEQUENCE DESCRIPTION: SEQ ID NO: 28:

US-09-066-046-28  
Query Match 8.5%; Score 71; DB 4; Length 3435;  
Best Local Similarity 58.3%; Pred. No. 2.6e-09;  
Matches 144; Conservative 0; Mismatches 100; Indels 3; Gaps 1;  
Qy 474 TGTAGTACTTAAATGACGGCATAACTTTTATGTCATTGATGGTTAATACTTGTCTATGA 533  
Db 904 TGAGGTGTTGAGATTAGGCGGTTTCTTCTACTTCTGTAATGGTCAATGCTTGTATGA 963  
Qy 534 CATTACAGCTGAAGGAGTATCTTTTCGTACCATATGTCATGTCAGGTATAGGACGATCT 593  
Db 964 TCTTCTTAGTGAAGGCTAGGCGGTGTTCTTACGCTTGTGTCGGTCTTGAGGTAACCT 1023  
Qy 594 TATCACTATTTTAAAGACC---TCATCTAAATTTGCTTACCAGGAAATAGGTAT 650  
Db 1024 CGTGGCGGTGTTGATGGGCATACCTCTTAAGCTTGTCTTATAGATTAAAGCGTGGGT 1083  
Qy 651 TAGTTACCCCTACACACAGAGTCTCTGCATTTTATGTTGATGATCTACCATGGGTTAT 710

Db 1084 GAGTATCAGCTCTCTCTGAAATCTCCGCTTTTGGGGGAGTCTTATCATCGCTGT 1143  
Qy 711 TGGTAAT 717  
Db 1144 GGGAGAT 1150

RESULT 35  
US-09-066-047-14  
; Sequence 14, Application US/09066047A  
; Patent No. 6306394  
; GENERAL INFORMATION:  
; APPLICANT: MURPHY, Cheryl  
; STOREY, James  
; BELTZ, Gerald A.  
; COUGHLIN, Richard T.  
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND METHODS OF  
; USE OF GRANULOCYTIC EHRlichIA  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HALE AND DORR LLP  
; STREET: 60 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: United States  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/066,047A  
; FILING DATE: 24-Apr-1998  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/044,869  
; FILING DATE: 25-APR-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Superko, Colleen  
; REGISTRATION NUMBER: 39,850  
; REFERENCE/DOCKET NUMBER: 106.941.156  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 526-6000  
; TELEFAX: (617) 526-6000  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3435 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
US-09-066-047-14

Query Match 8.5%; Score 71; DB 4; Length 3435;  
Best Local Similarity 58.3%; Pred. No. 2.6e-09;  
Matches 144; Conservative 0; Mismatches 100; Indels 3; Gaps 1;  
Qy 474 TGTAGTACTTAAATGACGGCATAACTTTTATGTCATTGATGGTTAATACTTGTCTATGA 533  
Db 904 TGAGGTGTTGAGATTAGGCGGTTTCTTCTACTTCTGTAATGGTCAATGCTTGTATGA 963  
Qy 534 CATTACAGCTGAAGGAGTATCTTTTCGTACCATATGTCATGTCAGGTATAGGACGATCT 593  
Db 964 TCTTCTTAGTGAAGGCTAGGCGGTGTTCTTACGCTTGTGTCGGTCTTGAGGTAACCT 1023  
Qy 594 TATCACTATTTTAAAGACC---TCATCTAAATTTGCTTACCAGGAAATAGGTAT 650  
Db 1024 CGTGGCGGTGTTGATGGGCATACCTCTTAAGCTTGTCTTATAGATTAAAGCGTGGGT 1083

Qy 651 TAGTTACCTATCACACAGAGTCTCTGCAATTTATGGTGGATACCTACCGGTTAT 710  
Db 1084 GAGTTATCAGCTCTCTCTGAAATCTCGCTTTTGTGGGGATTTCTATCATCGGTTGT 1143  
Qy 711 TGGAAT 717  
Db 1144 GGGAGAT 1150

RESULT 36  
US-09-066-046-28/c  
; Sequence 28, Application US/09066046A  
; Patent No. 6204252  
; GENERAL INFORMATION:  
; APPLICANT: MURPHY, Cheryl  
; STOREY, James  
; BELTZ, Gerald A.  
; COUGHLIN, Richard T.  
; TITLE OF INVENTION: CHARACTERIZATION OF GRANULOCYTIC  
; EHRLLICHIA AND METHODS OF USE

NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HALE AND DORR LLP  
STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: United States  
ZIP: 02109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/066,046A  
FILING DATE: 24-Apr-1998  
CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:  
NAME: Superko, Colleen  
REGISTRATION NUMBER: 39,850  
REFERENCE/DOCKET NUMBER: 106.941.155  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 526-6000  
TELEFAX: (617) 526-5000

INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3435 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

ANTI-SENSE: NO  
SEQUENCE DESCRIPTION: SEQ ID NO: 28:  
US-09-066-046-28

Query Match 7.1%; Score 60; DB 4; Length 3435;  
Best Local Similarity 54.5%; Pred. No. 1.8e-06;  
Matches 120; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

Qy 556 TTCGTACCATATGCATGTCAGGTATAGGACAGATCTTATCATTATTTAAAGACCTC 615  
Db 3409 TTCCCTCCTTACATGTGTGCTGGGTAGGGCGAGCTTTATTAATATCGCGATCACGTT 3350  
Qy 616 AATCTAAATTTGCTTACCAGGAAATAGTATTAGTTACCTTATCACACACAGAGTC 675  
Db 3349 ACAAGTAGTTGGCTTATAGAGGAAGTAGGCGTAAGTTACAGCTTACTCTCGAATA 3290  
Qy 676 TCTGCATTTATTTGGTGGTACTACCATGGCGTTATTTGGTAATAAATTTGAGAAGATACCT 735  
Db 3289 TCTTTAATAGCTGGAGGTTTTTACCACGGAATTTTCGATGACAGTATGCAGGTATTTCCC 3230  
Qy 736 GTAATAACTCTCTAGTATTAAATGATGCTCCTCAAAACCA 775

Db 3229 GCTAGTAACCGTGTAAACATAGCAGGTGGTGTGTCAGCCA 3190

RESULT 37  
US-09-066-047-14/c  
; Sequence 14, Application US/09066047A  
; Patent No. 6306394  
; GENERAL INFORMATION:  
; APPLICANT: MURPHY, Cheryl  
; STOREY, James  
; BELTZ, Gerald A.  
; COUGHLIN, Richard T.  
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND METHODS OF  
; USE OF GRANULOCYTIC EHRLLICHIA

NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HALE AND DORR LLP  
STREET: 60 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: United States  
ZIP: 02109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/066,047A  
FILING DATE: 24-Apr-1998  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/044,869  
FILING DATE: 25-Apr-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Superko, Colleen  
REGISTRATION NUMBER: 39,850  
REFERENCE/DOCKET NUMBER: 106.941.156  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 526-6000  
TELEFAX: (617) 526-5000

INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3435 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

ANTI-SENSE: NO  
SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
US-09-066-047-14

Query Match 7.1%; Score 60; DB 4; Length 3435;  
Best Local Similarity 54.5%; Pred. No. 1.8e-06;  
Matches 120; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

Qy 556 TTCGTACCATATGCATGTCAGGTATAGGACAGATCTTATCATTATTTAAAGACCTC 615  
Db 3409 TTCCCTCCTTACATGTGTGCTGGGTAGGGCGAGCTTTATTAATATCGCGATCACGTT 3350  
Qy 616 AATCTAAATTTGCTTACCAGGAAATAGTATTAGTTACCTTATCACACACAGAGTC 675  
Db 3349 ACAAGTAGTTGGCTTATAGAGGAAGTAGGCGTAAGTTACAGCTTACTCTCGAATA 3290  
Qy 676 TCTGCATTTATTTGGTGGTACTACCATGGCGTTATTTGGTAATAAATTTGAGAAGATACCT 735  
Db 3289 TCTTTAATAGCTGGAGGTTTTTACCACGGAATTTTCGATGACAGTATGCAGGTATTTCCC 3230  
Qy 736 GTAATAACTCTCTAGTATTAAATGATGCTCCTCAAAACCA 775  
Db 3229 GCTAGTAACCGTGTAAACATAGCAGGTGGTGTGTCAGCCA 3190

```

RESULT 38
US-09-288-339-7
; Sequence 7, Application US/09288339
; Patent No. 6436399
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Zhi, Ning
; APPLICANT: Onashi, No. 6436399iio
; TITLE OF INVENTION: Nucleic Acid Encoding the Major Outer Membrane Protein of
; TITLE OF INVENTION: The Causative Agent of Human Granulocytic Ehrlichiosis
; TITLE OF INVENTION: and Peptides Encoded Thereby
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/288,339
; CURRENT FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 837
; TYPE: DNA
; ORGANISM: P44-15
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(837)
US-09-288-339-7

```

Query Match	6.5%; Score 54.6; DB 4; Length 837;
Best Local Similarity	56.8%; Pred. No. 2.8e-05;
Matches 121; Conservative 0; Mismatches 89; Indels 3; Gaps 17;	
Qy.	474 TGTAGTACTTTAAAAATGACGGCATAACTTTTATGTGTCATTGATGGTTAATACTTGTCTATGA 533
Db	615 TGAGTTATTGAATTCAGGCGGGTTTCTTCTACTTCTGTGATGTCATGCTTGTATGA 674
Qy	534 CATTACAGCTGAAGGAGTATCTTTGTCACCATATGCATGTGCAGGTATAGGACGACAGTCT 593
Db	675 TCTTCTTAGTGAAGGTTTAGGTGTTGTCCTTAATGCTTGTGTGCTTCTGTGGGTAACTT 734
Qy	594 TATCACATATTTTAAAGACC--TCAATCTAAAAATTTGCTTTACCAAGGAAAAATAGGTAT 650
Db	735 CGTAGGTGTTGTTGATGGGCATATCACTCTTAAGCTTGCTTATAGATTAAAGGCTGGCTT 794
Qy	651 TAGTTACCTCATCACACAGAGTCTCTGCATT 683
Db	795 GAGTTATCAGCTCTCTCTGTGAATCTCTCTTT 827

RESULT 39  
 US-08-181-271A-36/c  
 ; Sequence 36, Application US/08181271A  
 ; Patent No. 5614395  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ryals, John A.  
 ; APPLICANT: Alexander, Danny C.  
 ; APPLICANT: Beck, James J.  
 ; APPLICANT: Duesing, John H.  
 ; APPLICANT: Friedrich, Leslie B.  
 ; APPLICANT: Goodman, Robert M.  
 ; APPLICANT: Harms, Christian  
 ; APPLICANT: Meins, Jr., Frederick  
 ; APPLICANT: Montoya, Alice  
 ; APPLICANT: Moyer, Mary B.  
 ; APPLICANT: Neuhaus, Jean-Marc  
 ; APPLICANT: Payne, George B.  
 ; APPLICANT: Sperison, Christoph.  
 ; APPLICANT: Stinson, Jeffrey R.  
 ; APPLICANT: Uknes, Scott J.  
 ; APPLICANT: Ward, Eric R.  
 ; APPLICANT: Williams, Shericea C.  
 ; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
 ; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 106

```

; ORGANISM: Cucumis sativus
; INDIVIDUAL ISOLATE: Cucumber Chitinase Genomic DNA
; IMMEDIATE SOURCE:
; CLONE: pBscucchrchts
US-08-181-271A-36

Query Match 6.0%; Score 50; DB 1; Length 12124;
Best Local Similarity 45.8%; Pred. No. 0.001;
Matches 173; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 169 GCTGAAGAACTCCTATTATGGAACAAATCTCTCAGTAAAGTTTCGGACTAAG 228
DB 11101 GGTCAAGATATCATATCAACATAGACATTTTATCTTAAAGAGTATATATAGAC 11042
QY 229 AAAGATGGTGATATACAAAAGAGACGATTTTACAGAGTAGTCTCCAGGCAATGATTTT 288
DB 11041 ATTAACATTTGATGAGAAGAAAAAATCAGAAAAATATGAGTTCTTGTGTTGGTGT 10982
QY 289 CAAATATCTTAATATCAGGATTTTCAGGAAGTATTTGTTACTCTATGGACGGACCAAGA 348
DB 10981 TATGTGAATATCATCTATTTGACTTAAACAACTAATTACATTTATGGAGAGATCAGTT 10922
QY 349 ATAGAATCTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGATAACAATGATACT 408
DB 10921 TTAATTACATNTAGCTATGTCACAGGTTGTTTTTTTATATCAAAATTAGAACATTAG 10862
QY 409 GATAATGCTGAATACTATAAAACATTTTGCATTTATCTCGTAAAGATGCAATGGAAGATCAG 468
DB 10861 GATAAAATTAACACTCTAATTAATCTATTTAGTTTAAATTCATCTGTATATATTATTA 10802
QY 469 CAATATGTAGTACTTAAATGACGGCATAACTTTTATGTCATTTGATGGTTAAATCTTGC 528
DB 10801 AGATTTTCAATTTTATATTTGATGGAATATATATATATATATATATATATATATTCGATC 10742
QY 529 TATGACATTACAGCTGAA 546
DB 10741 TATGACATGTTACTCAA 10724

RESULT 40
US-08-449-315-36/c
Sequence 36, Application US/08449315
Patent No. 5650505
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,315
FILING DATE: 24-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 12124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Cucumis sativus
INDIVIDUAL ISOLATE: Cucumber Chitinase Genomic DNA
IMMEDIATE SOURCE:
CLONE: pBscucchrchts
US-08-449-315-36

Query Match 6.0%; Score 50; DB 1; Length 12124;
```

Best Local Similarity 45.8%; Pred. No. 0.001;  
Matches 173; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

Qy	169	GCTGAAGAACTCCTATTATGGAACAAATCTCTCACTTAAAAAGTTTTTCGGACTAAAG	228
Db	11101	GGTCAAGAAATCATATCAACATAGACATTTTATACCTTAAAAAAGAGTAATAATAGAC	11042
Qy	229	AAAGATGGTGATATAACAACAAAAAGACGATTTTACAAGAGTAGCTCCAGGCATTGATTTT	288
Db	11041	ATTAACATTGATGAGAAGAAAAAATCAGAAAAATATGAGTTCTTTGTGTTTTGGTGT	10982
Qy	289	CAAAATAACTTAATATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGACGACCAAGA	348
Db	10981	TATGTGAATATCACTATTGACTTTAAAAACACAACTAATTACATTATGGAGAGATCAGTT	10922
Qy	349	ATAGAACCTTGAAGCTGCATATCAACAATTTAATCCAAAAAACCCGATAACCAATGATACT	408
Db	10921	TTATTTACTATTGAGCTATGTCACAAGGTTGTTTTTTTATAATCAAAATTAGAACATTAG	10862
Qy	409	GATAATGGTGATATATAACATTTTGCATTTATCTCGTAAAGATGCAATGGAAGATCAG	468
Db	10861	GATAAAATTAACAACTTCTAAATTAATTTAGTTTAAATTTTCATCTCTGTATATTATTTAA	10802
Qy	469	CAATATGTAGTACTTAAAAATGACGCGATAACTTTTTATGTCATTTGATGGTTAATACTTGC	528
Db	10801	AGATTTTCATTTTTTATATTGATGGAAATATATATATATATATATATATATATATGACATC	10742
Qy	529	TATGACATTACAGCTGAA	546
Db	10741	TATGACAATGGTACTCAA	10724

Search completed: July 4, 2003, 01:03:06  
Job time : 53 secs



Result No.	Query			Description		
	Score	Match	Length	ID	DB	
1	840	100.0	840	9	US-10-062-051-41	Sequence 41, Appl
2	840	100.0	840	9	US-10-062-920-41	Sequence 41, Appl
3	840	100.0	840	12	US-10-062-624-41	Sequence 41, Appl
4	840	100.0	843	9	US-10-314-639-47	Sequence 47, Appl
5	840	100.0	843	12	US-10-059-964-47	Sequence 47, Appl
6	598.6	71.3	852	9	US-10-314-639-3	Sequence 3, Appl
7	598.6	71.3	852	12	US-10-059-964-3	Sequence 3, Appl
8	238.4	28.4	924	9	US-10-314-639-35	Sequence 35, Appl
9	238.4	28.4	924	12	US-10-059-964-35	Sequence 35, Appl
10	238.4	28.4	1607	9	US-10-062-081-1	Sequence 1, Appl
11	238.4	28.4	1607	9	US-10-062-920-1	Sequence 1, Appl
12	238.4	28.4	1607	12	US-10-062-624-1	Sequence 1, Appl
13	234	27.9	840	9	US-10-062-051-5	Sequence 5, Appl
14	234	27.9	840	9	US-10-062-920-5	Sequence 5, Appl
15	234	27.9	840	12	US-10-062-624-5	Sequence 5, Appl
16	232.4	27.7	843	9	US-10-314-639-37	Sequence 37, Appl
17	232.4	27.7	843	12	US-10-059-964-37	Sequence 37, Appl
18	232.4	27.7	846	9	US-10-314-639-1	Sequence 1, Appl
19	232.4	27.7	846	12	US-10-059-964-1	Sequence 1, Appl

QY 181 CCTATTAAATGGAACAAATCTCTCACTAAAAAGTTTCGGACTAAAGAAAGATGGTGAT 240  
DB 181 CCTATTAAATGGAACAAATCTCTCACTAAAAAGTTTCGGACTAAAGAAAGATGGTGAT 240  
QY 241 ATAACAAAAAGAGACGATTTTACAAGAGTAGCTCCAGGCAATTTGATTTTCAAAATAACTTTA 300  
DB 241 ATAACAAAAAGAGACGATTTTACAAGAGTAGCTCCAGGCAATTTGATTTTCAAAATAACTTTA 300  
QY 301 ATATCAGGATTTTTCAGGAAGTATTTGGTTACTCTATGAGAGCGGACCAAGATAGAACTTGA 360  
DB 301 ATATCAGGATTTTTCAGGAAGTATTTGGTTACTCTATGAGAGCGGACCAAGATAGAACTTGA 360  
QY 361 GCTGCATATCAAAATTTTAAATCCAAAAACACCGATAACAATGATGATGATGATGATGATG 420  
DB 361 GCTGCATATCAAAATTTTAAATCCAAAAACACCGATAACAATGATGATGATGATGATGATG 420  
QY 421 TACTATAAAATTTTGCATTTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTAGTA 480  
DB 421 TACTATAAAATTTTGCATTTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTAGTA 480  
QY 481 CTTAAATGAGCGGATTAATTTTATGTCATTTGATGTTGATGTTGATGTTGATGTTGATG 540  
DB 481 CTTAAATGAGCGGATTAATTTTATGTCATTTGATGTTGATGTTGATGTTGATGTTGATG 540  
QY 541 GCTGAAGGAGTATCTTTCGTACCATATGATGTCAGGATATAGGAGCAGATCTTATCACT 600  
DB 541 GCTGAAGGAGTATCTTTCGTACCATATGATGTCAGGATATAGGAGCAGATCTTATCACT 600  
QY 601 ATTTTAAAGACCTCAATCTAAATTTTGTACCAAGGAAAAATAGGTATTTAGTTACCT 660  
DB 601 ATTTTAAAGACCTCAATCTAAATTTTGTACCAAGGAAAAATAGGTATTTAGTTACCT 660  
QY 661 ATACACACGAGAGTCTGCAATTTATGTCAGTATGATGATGATGATGATGATGATGATG 720  
DB 661 ATACACACGAGAGTCTGCAATTTATGTCAGTATGATGATGATGATGATGATGATGATG 720  
QY 721 TTTGAGAGATACCTGTAATACTCTCTGATGATGATGATGATGATGATGATGATGATGATG 780  
DB 721 TTTGAGAGATACCTGTAATACTCTCTGATGATGATGATGATGATGATGATGATGATGATG 780  
QY 781 GCTTCAGTAACCTTTGACGTTGGATCTTTGGCGGAGAAATTTGGAATGAGGTTTCACTTC 840  
DB 781 GCTTCAGTAACCTTTGACGTTGGATCTTTGGCGGAGAAATTTGGAATGAGGTTTCACTTC 840

## RESULT 2

US-10-062-920-41  
; Sequence 41, Application US/10062920  
; Publication No. US20030096250A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: US/10/062,920  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: US/09/660,587  
; PRIOR FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 41  
; LENGTH: 840  
; TYPE: DNA  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: nucleic acid sequence of E. canis p28-2

Query Match 100.0%; Score 840; DB 9; Length 840;  
Best Local Similarity 100.0%; Pred. No. 6.5e-169;

Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGAATTATAGAAAAATTTCTAGTAAGAGCGCGTTAATCTCATTTAATGTCAATCTTACCA 60  
DB 1 ATGAATTATAGAAAAATTTCTAGTAAGAGCGCGTTAATCTCATTTAATGTCAATCTTACCA 60  
QY 61 TATCAGTCTTTTTCAGAGATCCTGTAGGTTCAAGAACTAATGATTAACAAGAGGCTTCTAC 120  
DB 61 TATCAGTCTTTTTCAGAGATCCTGTAGGTTCAAGAACTAATGATTAACAAGAGGCTTCTAC 120  
QY 121 ATTAGTCAAAAGTCAATCCAAAGTATATCACACTTTTAGAAAAATTTCTCTGCTGAAGAACT 180  
DB 121 ATTAGTCAAAAGTCAATCCAAAGTATATCACACTTTTAGAAAAATTTCTCTGCTGAAGAACT 180  
QY 181 CCTATTAAATGGAACAAATTTCTCTCACTAAAAAGTTTCGGACTAAAGAAAGATGGTGAT 240  
DB 181 CCTATTAAATGGAACAAATTTCTCTCACTAAAAAGTTTCGGACTAAAGAAAGATGGTGAT 240  
QY 241 ATAACAAAAAGAGCGATTTTACAAAGATGATGTTTCTATGAGCGGACCAAGAAATAGAACTTGA 300  
DB 241 ATAACAAAAAGAGCGATTTTACAAAGATGATGTTTCTATGAGCGGACCAAGAAATAGAACTTGA 300  
QY 301 ATATCAGGATTTTTCAGGAAGTATTTGGTTACTCTATGAGCGGACCAAGAAATAGAACTTGA 360  
DB 301 ATATCAGGATTTTTCAGGAAGTATTTGGTTACTCTATGAGCGGACCAAGAAATAGAACTTGA 360  
QY 361 GCTGCATATCAAAATTTTATCCAAAAACACCGATAACAATGATGATGATGATGATGATGATG 420  
DB 361 GCTGCATATCAAAATTTTATCCAAAAACACCGATAACAATGATGATGATGATGATGATGATG 420  
QY 421 TACTATAAAATTTTGCATTTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTAGTA 480  
DB 421 TACTATAAAATTTTGCATTTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTAGTA 480  
QY 481 CTTAAATGAGCGGATTAATTTTATGTCATTTGATGTTGATGTTGATGTTGATGTTGATG 540  
DB 481 CTTAAATGAGCGGATTAATTTTATGTCATTTGATGTTGATGTTGATGTTGATGTTGATG 540  
QY 541 GCTGAAGGAGTATCTTTCGTACCATATGATGTCAGGATATAGGAGCAGATCTTATCACT 600  
DB 541 GCTGAAGGAGTATCTTTCGTACCATATGATGTCAGGATATAGGAGCAGATCTTATCACT 600  
QY 601 ATTTTAAAGACCTCAATCTAAATTTTGTACCAAGGAAAAATAGGTATTTAGTTACCT 660  
DB 601 ATTTTAAAGACCTCAATCTAAATTTTGTACCAAGGAAAAATAGGTATTTAGTTACCT 660  
QY 661 ATACACACGAGAGTCTGCAATTTATGTCAGTATGATGATGATGATGATGATGATGATG 720  
DB 661 ATACACACGAGAGTCTGCAATTTATGTCAGTATGATGATGATGATGATGATGATGATG 720  
QY 721 TTTGAGAGATACCTGTAATACTCTCTGATGATGATGATGATGATGATGATGATGATGATG 780  
DB 721 TTTGAGAGATACCTGTAATACTCTCTGATGATGATGATGATGATGATGATGATGATGATG 780  
QY 781 GCTTCAGTAACCTTTGACGTTGGATCTTTGGCGGAGAAATTTGGAATGAGGTTTCACTTC 840  
DB 781 GCTTCAGTAACCTTTGACGTTGGATCTTTGGCGGAGAAATTTGGAATGAGGTTTCACTTC 840

## RESULT 3

US-10-062-624-41  
; Sequence 41, Application US/10062624  
; Patent No. US20020115840A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-Kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2/D1  
; CURRENT APPLICATION NUMBER: US/10/062,624  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: 09/660,587

```
; PRIOR FILING DATE: 2000-09-12
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 41
; LENGTH: 840
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: nucleic acid sequence of E. canis p28-2
US-10-062-624-41

Query Match      100.0%; Score 840; DB 12; Length 840;
Best Local Similarity 100.0%; Pred. No. 6.5e-169;
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAATTATAAGAAAATCTAGTAAAGACGCGTTAAATCTCATTAAATGCTCAATCTTACCA 60
Db 1 ATGAATTATAAGAAAATCTAGTAAAGACGCGTTAAATCTCATTAAATGCTCAATCTTACCA 60

Qy 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120
Db 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120

Qy 121 ATTAGTCAAAAGTACAATCCAAAGTATATACACTTTTGAAGAAATCTCTGCTGAAGAACT 180
Db 121 ATTAGTCAAAAGTACAATCCAAAGTATATACACTTTTGAAGAAATCTCTGCTGAAGAACT 180

Qy 181 CCTATTAATGGAACAAATCTCTCACTAAAGAAAGTTTTCGAGCTAAAGAAAGATGGTGAT 240
Db 181 CCTATTAATGGAACAAATCTCTCACTAAAGAAAGTTTTCGAGCTAAAGAAAGATGGTGAT 240

Qy 241 ATACAAAAAAGACGATTTTCAAGAGTAGTCCAGGCATGATTTTCAAAAATAACTTTA 300
Db 241 ATACAAAAAAGACGATTTTCAAGAGTAGTCCAGGCATGATTTTCAAAAATAACTTTA 300

Qy 301 ATATCAGGATTTTTCAGGAAGTATGGTTACTCTATGGAACGACCAAGAAATAGAACTTGA 360
Db 301 ATATCAGGATTTTTCAGGAAGTATGGTTACTCTATGGAACGACCAAGAAATAGAACTTGA 360

Qy 361 GCTGCATATCAACAAATTTTAAATCCAAAAACCCGATATCAATGATGATCTGATGATGTA 420
Db 361 GCTGCATATCAACAAATTTTAAATCCAAAAACCCGATATCAATGATGATCTGATGATGTA 420

Qy 421 TACTATAAACAATTTTGCATTTATGTCATTTGATGTTTAAATGCTATGATGATGATGTA 480
Db 421 TACTATAAACAATTTTGCATTTATGTCATTTGATGTTTAAATGCTATGATGATGATGTA 480

Qy 481 CTTAAAAATGACGCAATAAATTTTATGTCATTTGATGTTTAAATGCTATGATGATGATGTA 540
Db 481 CTTAAAAATGACGCAATAAATTTTATGTCATTTGATGTTTAAATGCTATGATGATGATGTA 540

Qy 541 GCTGAAGAGTATCTTTGCTACCATATGCTGAGTATGCTGAGTATGAGGAGATCTTATCACT 600
Db 541 GCTGAAGAGTATCTTTGCTACCATATGCTGAGTATGCTGAGTATGAGGAGATCTTATCACT 600

Qy 601 ATTTTAAAGACCTCAATCTAAATTTTGCCTTACCAAGGAAAAATAGGATTTAGTTACCCCT 660
Db 601 ATTTTAAAGACCTCAATCTAAATTTTGCCTTACCAAGGAAAAATAGGATTTAGTTACCCCT 660

Qy 661 ATCCACCAAGTCTCTGCAATTTTATGTCGATGATGATGATGATGATGATGATGATGATG 720
Db 661 ATCCACCAAGTCTCTGCAATTTTATGTCGATGATGATGATGATGATGATGATGATGATG 720

Qy 721 TTTGAGAGATACCTGTAAATCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 780
Db 721 TTTGAGAGATACCTGTAAATCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 780

Qy 781 GCTTCAGTAACTTTGAGGTTGGATGATCTTTGGCGAGAAATTTGGAATGAGGTTTCACTTTC 840
Db 781 GCTTCAGTAACTTTGAGGTTGGATGATCTTTGGCGAGAAATTTGGAATGAGGTTTCACTTTC 840

RESULT 4
US-10-314-639-47

; Sequence 47, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Ohasi, No. US20030103991A1
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; PRIOR FILING DATE: 2002-12-09
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 47
; LENGTH: 843
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(843)
US-10-314-639-47

Query Match      100.0%; Score 840; DB 9; Length 843;
Best Local Similarity 100.0%; Pred. No. 6.5e-169;
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAATTATAAGAAAATCTAGTAAAGACGCGTTAAATCTCATTAAATGCTCAATCTTACCA 60
Db 1 ATGAATTATAAGAAAATCTAGTAAAGACGCGTTAAATCTCATTAAATGCTCAATCTTACCA 60

Qy 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120
Db 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120

Qy 121 ATTAGTCAAAAGTACAATCCAAAGTATATACACTTTTGAAGAAATCTCTGCTGAAGAACT 180
Db 121 ATTAGTCAAAAGTACAATCCAAAGTATATACACTTTTGAAGAAATCTCTGCTGAAGAACT 180

Qy 181 CCTATTAATGGAACAAATCTCTCACTAAAGAAAGTTTTCGAGCTAAAGAAAGATGGTGAT 240
Db 181 CCTATTAATGGAACAAATCTCTCACTAAAGAAAGTTTTCGAGCTAAAGAAAGATGGTGAT 240

Qy 241 ATACAAAAAAGACGATTTTCAAGAGTAGTCCAGGCATGATTTTCAAAAATAACTTTA 300
Db 241 ATACAAAAAAGACGATTTTCAAGAGTAGTCCAGGCATGATTTTCAAAAATAACTTTA 300

Qy 301 ATATCAGGATTTTTCAGGAAGTATGGTTACTCTATGGAACGACCAAGAAATAGAACTTGA 360
Db 301 ATATCAGGATTTTTCAGGAAGTATGGTTACTCTATGGAACGACCAAGAAATAGAACTTGA 360

Qy 361 GCTGCATATCAACAAATTTTAAATCCAAAAACCCGATATCAATGATGATCTGATGATGTA 420
Db 361 GCTGCATATCAACAAATTTTAAATCCAAAAACCCGATATCAATGATGATCTGATGATGTA 420

Qy 421 TACTATAAACAATTTTGCATTTATGTCATTTGATGTTTAAATGCTATGATGATGATGTA 480
Db 421 TACTATAAACAATTTTGCATTTATGTCATTTGATGTTTAAATGCTATGATGATGATGTA 480

Qy 481 CTTAAAAATGACGCAATAAATTTTATGTCATTTGATGTTTAAATGCTATGATGATGATGTA 540
Db 481 CTTAAAAATGACGCAATAAATTTTATGTCATTTGATGTTTAAATGCTATGATGATGATGTA 540

Qy 541 GCTGAAGAGTATCTTTGCTACCATATGCTGAGTATGCTGAGTATGAGGAGATCTTATCACT 600
Db 541 GCTGAAGAGTATCTTTGCTACCATATGCTGAGTATGCTGAGTATGAGGAGATCTTATCACT 600

Qy 601 ATTTTAAAGACCTCAATCTAAATTTTGCCTTACCAAGGAAAAATAGGATTTAGTTACCCCT 660
Db 601 ATTTTAAAGACCTCAATCTAAATTTTGCCTTACCAAGGAAAAATAGGATTTAGTTACCCCT 660

Qy 661 ATCCACCAAGTCTCTGCAATTTTATGTCGATGATGATGATGATGATGATGATGATGATG 720
Db 661 ATCCACCAAGTCTCTGCAATTTTATGTCGATGATGATGATGATGATGATGATGATGATG 720

Qy 721 TTTGAGAGATACCTGTAAATCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 780
Db 721 TTTGAGAGATACCTGTAAATCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 780

Qy 781 GCTTCAGTAACTTTGAGGTTGGATGATCTTTGGCGAGAAATTTGGAATGAGGTTTCACTTTC 840
Db 781 GCTTCAGTAACTTTGAGGTTGGATGATCTTTGGCGAGAAATTTGGAATGAGGTTTCACTTTC 840

RESULT 4
US-10-314-639-47
```

Db 661 ATCACACAGAGTCTCTGCATTTATTTGGTGATACCTACCTACCTGCGGTTATTTGGTAATAAA 720  
QY 721 TTTGAGAGATACCTGTGTAATAAATCTCTGTAGTATTAATAATGATGCTCCTCAAAACCAATCT 780  
Db 721 TTTGAGAGATACCTGTGTAATAAATCTCTGTAGTATTAATAATGATGCTCCTCAAAACCAATCT 780  
QY 781 GCTTCAGTAACCTCTGAGCTTGGATCTTTGGCGAGAAATTTGGAAATGAGGTTTCACTTC 840  
Db 781 GCTTCAGTAACCTCTGAGCTTGGATCTTTGGCGAGAAATTTGGAAATGAGGTTTCACTTC 840

## RESULT 5

US-10-059-964-47  
; Sequence 47, Application US/10059964  
; Patent No. US20020120115A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20020120115A1io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; TITLE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/059,964  
; EARLIER FILING DATE: 2002-01-28  
; EARLIER FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 47  
; LENGTH: 843  
; TYPE: DNA  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(843)  
US-10-059-964-47

Query Match 100.0%; Score 840; DB 12; Length 843;  
Best Local Similarity 100.0%; Pred. No. 6.5e-169; Indels 0; Gaps 0;  
Matches 840; Conservative 0; Mismatches 0

QY 1 ATGAATTAAGAAAATCTAGTAAGAGCGGTTAATCTCAATTAATGTCAATCTTACCA 60  
Db 1 ATGAATTAAGAAAATCTAGTAAGAGCGGTTAATCTCAATTAATGTCAATCTTACCA 60  
QY 61 TATCAGTCTTTTGCAGATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 120  
Db 61 TATCAGTCTTTTGCAGATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 120  
QY 121 ATTAGTGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180  
Db 121 ATTAGTGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180  
QY 181 CCTATTATGGAACAAATCTCTCACTAAAGAAATTTTCGGACTAAAGAAAGATGATGAT 240  
Db 181 CCTATTATGGAACAAATCTCTCACTAAAGAAATTTTCGGACTAAAGAAAGATGATGAT 240  
QY 241 ATACAAAGAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
Db 241 ATACAAAGAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
QY 301 ATATCAGGATTTTTCAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
Db 301 ATATCAGGATTTTTCAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
QY 361 GCTGCATATCAACAAATTTAATCCAAAAACACCGATACAAATGATGATGATGATGATGATGAT 420  
Db 361 GCTGCATATCAACAAATTTAATCCAAAAACACCGATACAAATGATGATGATGATGATGATGAT 420  
QY 421 TACTATAACATTTTGCATATCTCTGTAAGATGATGATGATGATGATGATGATGATGATGATGAT 480  
Db 421 TACTATAACATTTTGCATATCTCTGTAAGATGATGATGATGATGATGATGATGATGATGATGAT 480  
QY 481 CTTAAAAATGACGGCATACCTTTTATGTCATTTGATGATGATGATGATGATGATGATGATGAT 540

Db 481 CTTAAAAATGACGGCATACCTTTTATGTCATTTGATGATGATGATGATGATGATGATGATGAT 540  
QY 541 GCTGAAGAGATCTTTTGCATACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
Db 541 GCTGAAGAGATCTTTTGCATACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
QY 601 ATTTTAAAGACCTCAATCTAAATTTTGTCTTACCAAGGAAATAGGTATAGTTACCTT 660  
Db 601 ATTTTAAAGACCTCAATCTAAATTTTGTCTTACCAAGGAAATAGGTATAGTTACCTT 660  
QY 661 ATCACACAGAGTCTCTGCATTTATTTGGTGATACCTACCTACCTGCGGTTATTTGGTAATAAA 720  
Db 661 ATCACACAGAGTCTCTGCATTTATTTGGTGATACCTACCTACCTGCGGTTATTTGGTAATAAA 720  
QY 721 TTTGAGAAGATACCTGTGATTAATCTCTGTAGTATTAATAATGATGATGATGATGATGATGAT 780  
Db 721 TTTGAGAAGATACCTGTGATTAATCTCTGTAGTATTAATAATGATGATGATGATGATGATGAT 780  
QY 781 GCTTCAGTAACCTCTGAGCTTGGATCTTTGGCGAGAAATTTGGAAATGAGGTTTCACTTC 840  
Db 781 GCTTCAGTAACCTCTGAGCTTGGATCTTTGGCGAGAAATTTGGAAATGAGGTTTCACTTC 840

## RESULT 6

US-10-314-639-3  
; Sequence 3, Application US/10314639  
; Publication No. US20030103991A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20030103991A1io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; TITLE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/314,639  
; PRIOR FILING DATE: 2002-12-09  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 852  
; TYPE: DNA  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(852)  
US-10-314-639-3

Query Match 71.3%; Score 598.6; DB 9; Length 852;  
Best Local Similarity 82.6%; Pred. No. 1.1e-117; Indels 9; Gaps 1;  
Matches 701; Conservative 0; Mismatches 139

QY 1 ATGAATTAAGAAAATCTAGTAAGAGCGGTTAATCTCAATTAATGTCAATCTTACCA 60  
Db 1 ATGAATTAAGAAAATCTAGTAAGAGCGGTTAATCTCAATTAATGTCAATCTTACCT 60  
QY 61 TATCAGTCTTTTGCAGATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 111  
Db 61 TACCAATCTTTTGCAGATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 120  
QY 112 GCTTCTTACATTTAGTGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 171  
Db 121 GCTTCTTACATTTAGTGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180  
QY 172 GAAGAACTCTCTATTATTAAGAAACAAATCTCTCACTAAAAAGTTTTCGGACTAAAGAAA 231  
Db 181 GAAGAGCTCCCATCATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
QY 232 GAT 291  
Db 241 GACGAGATATAGCAATCTCGCAATCTCGCAATCTCGCAATCTCGCAATCTCGCAATCTCGCA 300

QY 292 AATACCTTAATATCAGGATTTTCAGGAAGTATGGTTACTCTATGGACGGACCAAGATA 351  
 DB |||||  
 QY 301 AATAACCTTAATATCAGGATTTTCAGGAAGTATGGTTACTCTATGGATGGGCCAAGATA 360  
 DB |||||  
 QY 352 GAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGATAACAATGACTGAT 411  
 DB |||||  
 QY 361 GAACTTGAAGCTGCATACCAAAAATTTGATGCAAAAAATCTCGACAACATGACACTAAT 420  
 DB |||||  
 QY 412 AATGGTGAATACTATAAACAATTTTGCATTTCTCGTAAAGATGCAATGGAAGATCAGCAA 471  
 DB |||||  
 QY 421 AGCGGTGACTACTATAAATACCTTTGGACTCTCGTGAAGACGCAATAGCAGATAAGAAA 480  
 DB |||||  
 QY 472 TATGTAGTACTATAAATGACGGCATACCTTTTATGTCATTTGATGGTTAATGCTGCTAT 531  
 DB |||||  
 QY 481 TATGTGTGCTTAAAAATGAAGGCATCACTTTTATGTCATTTAATGGTTAAACACTTGTCTAT 540  
 DB |||||  
 QY 532 GACATTCAGCTGAAGGAGTATCTTTTCGTACCATATGCTGCGAGGTATAGGAGCAGAT 591  
 DB |||||  
 QY 541 GACATTCAGCTGAAGGAGTATCTTTTCATCCGATATGCTGCAAGGTGTAGGAGCAGAC 600  
 DB |||||  
 QY 592 CTTATACCTATTTTAAAGACCTCAATCTAAATTTTGCCTTACCAAGGAAAAATAGGTATT 651  
 DB |||||  
 QY 601 CTTATAAAGCTATTTAAGGATTTTAAATTTAAATTTCTCATACCAAGGAAAAATAGGTATT 660  
 DB |||||  
 QY 652 AGTTACCTATCACACGAGATCTCTGCATTTTATGTTGATGATACCTACCTAGCGGTATT 711  
 DB |||||  
 QY 661 AGCTATCCAAATCACACGAGATTTTCGCTTTTATTTGAGGATACCTACCAAGGAGTTATA 720  
 DB |||||  
 QY 712 GGTAAATTTTGAAGATACCTGTAAATACTCTGTAGTATTAATGATGCTCCCTCAA 771  
 DB |||||  
 QY 721 GGAATAATTTTAAACAAATACCTGTAAATAACCTGTAGTATTTAGAGGAGCTCCCTCAA 780  
 DB |||||  
 QY 772 ACCACATCTGCTCAGTAACCTTTGACGTTGGATATCTTTGGCGGAGAAATTTGGAATGAGG 831  
 DB |||||  
 QY 781 ACAACATCTGGCTAGTAACCTATTGACACTGGATCTTTGGCGGAGAAATTTGGAATGAGG 840  
 DB |||||  
 QY 832 TTCACCTTC 840  
 DB |||||  
 QY 841 TTCACCTTC 849

## RESULT 7

US-10-059-964-3  
 ; Sequence 3, Application US/10059964  
 ; Patent No. US20020120115A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rikihisa, Yasuko  
 ; APPLICANT: Ohasi, No. US20020120115A1io  
 ; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
 ; FILE REFERENCE: Chaffeensis  
 ; FILE REFERENCE: 22727/04021  
 ; CURRENT APPLICATION NUMBER: US/10/059,964  
 ; CURRENT FILING DATE: 2002-01-28  
 ; EARLIER APPLICATION NUMBER: 09/314,701  
 ; EARLIER FILING DATE: 1999-05-19  
 ; NUMBER OF SEQ ID NOS: 66  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 3  
 ; LENGTH: 852  
 ; TYPE: DNA  
 ; ORGANISM: Ehrlichia chaffeensis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(852)  
 US-10-059-964-3

Query Match 71.3%; Score 598.6; DB 12; Length 852;

Best Local Similarity 82.6%; Pred. No. 1.1e-117;

Matches 701; Conservative 0; Mismatches 139; Indels 9; Gaps 1;

QY 1 ATGAATTATAGAAATCTAGTAAAGCGCGTTAATCTCATTAATGTCATCTTACCA 60

DB 1 ATGAATTATAGAAATCTAGTAAAGCGCGTTAATCTCATTAATGTCATCTTACCT 60

QY 61 TATCAGTCTTTTGCAGATCCTGTAGGTTCAA-----GAACTAATGATAACAAAGAA 111  
 DB |||||  
 QY 61 TACCAATCTTTTGCAGATCCTGTAACTTTCAATGATACAGGAATCAACGACAGCAGAGAA 120  
 DB |||||  
 QY 112 GGCTTCTACATTAGTGCAGAGTACAAATCCAAGTATATCACACTTTAGAAAATTTCTGCT 171  
 DB |||||  
 QY 121 GGCTTCTACATTAGTGTAAAGTATAATCCAAGCATATCACACTTTAGAAAATTTCTGCT 180  
 DB |||||  
 QY 172 GAAGAACTCTCTAATTAATGGAACAAAATTTCTCTCAATAAAAAAGTTTTCGAGCTAAAGAAA 231  
 DB |||||  
 QY 181 GAAGAACTCTCTAATTAATGGAACAAAATTTCTCTCAATAAAAAAGTTTTCGAGCTAAAGAAA 240  
 DB |||||  
 QY 232 GATGGTGAATATAACAAAAAAGACGATTTTACAAGAGTAGCTCCAGGCATTTGATTTTCAA 291  
 DB |||||  
 QY 241 GACGAGATATAGCAAACTCTCGAAATTTTAAACAGGACAGATCCAGCCCTCGAGTTTCAG 300  
 DB |||||  
 QY 292 AATTAACCTTAATATCAGGATTTTCAGGAAGTATTTGGTTACTCTATGGACGGACCAAGATA 351  
 DB |||||  
 QY 301 AATAACCTTAATATCAGGATTTTCAGGAAGTATTTGGTTACTCTATGGATGGGCCAAGATA 360  
 DB |||||  
 QY 352 GAACTTGAAGCTGCATATCAACAATTTTAAATCCAAAAACACCGATAACAATGACTGAT 411  
 DB |||||  
 QY 361 GAACTTGAAGCTGCATACCAAAAATTTGATGCAAAAAATCTCGACAACATGACACTAAT 420  
 DB |||||  
 QY 412 AATGGTGAATATAAACAATTTTGCATTTCTCGTAAAGATGCAATGGAAGATCAGCAA 471  
 DB |||||  
 QY 421 AGCGGTGACTACTATAAATACCTTTTGGACTATCTCGTGAAGACGCAATAGCAGATAAGAAA 480  
 DB |||||  
 QY 472 TATGTAGTACTATAAATGACGGCATACCTTTTATGTCATTTGATGGTTAATGCTGCTAT 531  
 DB |||||  
 QY 481 TATGTGTGCTTAAAAATGAAGGCATCACTTTTATGTCATTTAATGGTTAAACACTTGTCTAT 540  
 DB |||||  
 QY 532 GACATTCAGCTGAAGGAGTATCTTTTCGTACCATATGCTGCGAGGTATAGGAGCAGAT 591  
 DB |||||  
 QY 541 GACATTCAGCTGAAGGAGTATCTTTTCATCCGATATGCTGCAAGGTGTAGGAGCAGAC 600  
 DB |||||  
 QY 592 CTTATACCTATTTTAAAGACCTCAATCTAAATTTTGCCTTACCAAGGAAAAATAGGTATT 651  
 DB |||||  
 QY 601 CTTATAAAGCTATTTAAGGATTTTAAATTTAAATTTCTCATACCAAGGAAAAATAGGTATT 660  
 DB |||||  
 QY 652 AGTTACCTATCACACGAGATCTCTGCATTTTATGTTGATGATACCTACCTAGCGGTATT 711  
 DB |||||  
 QY 661 AGCTATCCAAATCACACGAGATTTTCGCTTTTATTTGAGGATACCTACCAAGGAGTTATA 720  
 DB |||||  
 QY 712 GGTAAATTTTGAAGATACCTGTAAATACTCTGTAGTATTAATGATGCTCCCTCAA 771  
 DB |||||  
 QY 721 GGAATAATTTTAAACAAATACCTGTAAATAACCTGTAGTATTTAGAGGAGCTCCCTCAA 780  
 DB |||||  
 QY 772 ACCACATCTGCTCAGTAACCTTTGACGTTGGATATCTTTGGCGGAGAAATTTGGAATGAGG 831  
 DB |||||  
 QY 781 ACAACATCTGGCTAGTAACCTATTGACACTGGATCTTTGGCGGAGAAATTTGGAATGAGG 840  
 DB |||||  
 QY 832 TTCACCTTC 840  
 DB |||||  
 QY 841 TTCACCTTC 849

## RESULT 8

US-10-314-639-35  
 ; Sequence 35, Application US/10314639  
 ; Publication No. US20030103991A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rikihisa, Yasuko  
 ; APPLICANT: Ohasi, No. US20030103991A1io  
 ; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
 ; FILE REFERENCE: Chaffeensis  
 ; FILE REFERENCE: 22727/04021  
 ; CURRENT APPLICATION NUMBER: US/10/314,639  
 ; CURRENT FILING DATE: 2002-12-09  
 ; PRIOR APPLICATION NUMBER: US/09/314,701  
 ; PRIOR FILING DATE: 1999-05-19  
 ; NUMBER OF SEQ ID NOS: 66



Db 680 CTATGTTGAAGTACAAAGTCTTAAATTTCTTACCAGGAAACTGGGCAATAGTTACT 739  
Qy 659 CTATCACACAGAGTCTCTGCATTTATGTGGTGATACATCGGCTTATGGTAAATA 718  
Db 740 CTATTAATCCGAAACCTCTGTTTCATCGGTGGCAATTTCCACAGGATCATAGTAAAG 799  
Qy 719 AATTGGAGAGATACCTGTATTAATCTCTGTAGTATTAATGATGCTCTCAACACCAT 778  
Db 800 AGTTTAGAGATATTCCTGCAATAGTACCTAGTAATCAATACAAAGTGGACCAAT 859  
Qy 779 CTGCTTCAGTAATCTCTGACCTTGGATCTTTGGCGAGAAATTTGGATGAGGTTCACT 838  
Db 860 TTGCAACAGTAACACTAAATGTGTCTACTTTGGTTTGAACCTTGGAGGAAGATTTAACT 919  
Qy 839 TC 840  
Db 920 TC 921

RESULT 10  
US-10-062-051-1  
; Sequence 1, Application US/10062051  
; Publication No. US20030073095A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: US/10/062,051  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: US/09/660,587  
; PRIOR FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 1  
; LENGTH: 1607  
; TYPE: DNA  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: nucleic acid sequence of E. canis p28-7

US-10-062-051-1  
Query Match 28.4%; Score 238.4; DB 9; Length 1607;  
Best Local Similarity 57.7%; Pred. No. 3.6e-41;  
Matches 486; Conservative 0; Mismatches 346; Indels 10; Gaps 3;  
Qy 1 ATGAATTATAGAAAAATCTAGTAAGAGCGGTTAATCTCATTAATGTCAATCTTACCA 60  
Db 146 ATGAATTGCAAAAAATCTTATAACAACTGCAATTAATCATTAATGTACTCTATTCCA 205  
Qy 61 TATCAGTCTTTTTCAGATCTCTGATGTTCAAGACTAATGATAACAAGAGGCTTCTAC 120  
Db 206 AGCATATCTTTTCTGATCTATCAAGATGTTAATGATGTTGTTAA-----CTTCTAT 259  
Qy 121 ATTAGTCAAGATCAATCAAGTATATCACTTTTAGAAAAATCTCTGCTGAAGA--AA 178  
Db 260 ATTAGTGAAGATATGACCAAGTGTCTCAATTTTGGTAGCTTCTCAGCTTAAGAAGAA 319  
Qy 179 CTCCTATTATGGAACAAATCTCTCACTAAAAAGTTTTCGGACTAAAGAAAGATGGTG 238  
Db 320 AGCAAAATCAACTGTTGGAGTTTGGATTAAACATGATTTGGATGGAGTCCAACTACT- 378  
Qy 239 ATATAACAAAAAGAGATTTTACAGAGTAGCTCAGGATTTGATTTTCAAAATTAAT 298  
Db 379 -TAAGAATAACACGCTGACTTTTACTGTTCCAAACTATTGTTCCAGATACGAGAACATC 437  
Qy 299 TAATATCAGGATTTTTCAGGAAGTATGTTGTTACTCTATGGACGGACCAAGATGAACCTG 358  
Db 438 CATTCTAGGGTTTGCAGGAGCTATCGTTACTCAATGGGTGGCCCCAAGAAGATTAATCG 497

Qy 359 AAGTCGCATATCAACAATTTAAATCCAAAAAACCCGATAACAATGATGATGATTAATGGTG 418  
Db 498 AAATATCTTATGAAGCATTCGACGTAAAGAGTCTTAATATCAATATATCAAAATGACGCGC 557  
Qy 419 AATACTATAAACATTTTGCATTTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTAG 478  
Db 558 ACAGGTACTGGCTCTATCTCATCCACATCGGCGAGCCATGGAAGCTGATAAATTTGTCT 617  
Qy 479 TACTTAAAAATGACGGCATAACTTTTATGTCAITGATGGTTAAATCTTGTCTATGACATTA 538  
Db 618 TCTTAAAAAACGAGGGTAAATTCACATATCACCTTGCATTAATGATGATGATATAA 677  
Qy 539 CAGCTGAAGAGATCTTTTCGTACCATATGATGCTGAGGTATAGGAGCAGATCTTATCA 598  
Db 678 TAAATGACAAAGTACCTGTTTCTCTTATATATGCGCAGGTATTTGGTACTGATTTGATT 737  
Qy 599 CTATTTTAAAGACCTCAATCTAAATTTCTTACCAGGAAATTAAGGTATTTAGTTTACC 658  
Db 738 CTATGTTTGAAGCTACAAGTCTTAAATTTCTTACCAGGAAATTAAGGTATTTAGTTTACT 797  
Qy 659 CTATCACACAGAGTCTCTGCATTTTATGTGTGATACTACCATGGGCTTATTTGGTAAATA 718  
Db 798 CTATTAATCCGAAACCTCTGTTTTCATCGTGGGCAATTTCCACAGGATCATAGTAAAG 857  
Qy 719 AATTGAGAGATACCTGTGTAATACTCTCTGTAGTATTAATGATGCTCTCAAAACCAAT 778  
Db 858 AGTTTAGAGATATTTCTGCAATAGTACTAGTAATCACTCACTACAATAAGTGGACCAAT 917  
Qy 779 CTGCTTCAGTAATCTGTAGTGGTGGTACTTTGGCGAGAAATTTGGAATGAGGTTTCACT 838  
Db 918 TTGCAACAGTAACACTAAATGTGTGTCACCTTTGGTTTAGAACTTTGGAGGAAGATTTAACT 977  
Qy 839 TC 840  
Db 978 TC 979

RESULT 11  
US-10-062-920-1  
; Sequence 1, Application US/10062920  
; Publication No. US20030096250A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: US/10/062,920  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: US/09/660,587  
; PRIOR FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 1  
; LENGTH: 1607  
; TYPE: DNA  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: nucleic acid sequence of E. canis p28-7

US-10-062-920-1  
Query Match 28.4%; Score 238.4; DB 9; Length 1607;  
Best Local Similarity 57.7%; Pred. No. 3.6e-41;  
Matches 486; Conservative 0; Mismatches 346; Indels 10; Gaps 3;  
Qy 1 ATGAATTATAGAAAAATCTAGTAAGAGCGGTTAATCTCATTAATGTCAATCTTACCA 60  
Db 146 ATGAATTGCAAAAAATCTTATAACAACTGCAATTAATCATTAATGTACTCTATTCCA 205  
Qy 61 TATCAGTCTTTTTCAGATCTCTGATGTTCAAGACTAATGATAACAAGAGGCTTCTAC 120

```
Db 206 AGCATATCTTTCTGATACATATCAAGATGTAACATGGGTGGTAA-----CTTCTAT 259
Qy 121 ATTAGTGAAGTAAATCCCAAGTATATCAACATTTAGAAAATCTCTGCTGAAGA--AA 178
Db 260 ATTAGTGAAGTATGTACCAAGTGTCTCACATTTTGGTAGCTTCTCAGCTAAAGAAGAA 319
Qy 179 CTCCTATTAAATGAACAATCTCTCACTAAATAAGTTTTCGGACTAAAGAAGATGGTG 238
Db 320 AGCAATCAACTGTTGGAGTTTTCGGATTAAACATGATTTGGATGGGAAGTCCAACTAT- 378
Qy 239 ATATAACAAAAAGACGATTTTCAAGAGTAGTCCAGGCATTCATTTTCAAAATAACT 298
Db 379 -TAAGAATAACACGCTGACTTACTGTTCCAACTATTGTTTCAGATACGGAACAATC 437
Qy 299 TAATATCAGGATTTTTCAGGAAGTATTTGTTACTCTATGGAACGGAACAAGATAGAACTTG 358
Db 438 CATTTCTAGGTTTTCAGGAGCTATCGGTTTACTCAATGGTGGCCCAAGAATAAGAAATTCG 497
Qy 359 AAGCTGCATATCAACAATTTAATCCAAAAACCCGATACCAATGATATGATATGATGGTG 418
Db 498 AATATCTTTATGAAGCATTCGACGTAAAGATGCTTAATATCAATATCAAAATGACGGC 557
Qy 419 AATACTATAAACATTTTTCGATATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTAG 478
Db 558 ACAGTACTCGCTCTATCTCATACACATCGGCAGCCATGGAAGCTGATAAATTTGTCT 617
Qy 479 TACTTAAAAATGACGGCATAACTTTTATGCTAGTATTAATGATCTCTCAAAACAAT 538
Db 618 TCTTAAAAACGAAGGTTAATTCGACATATCACTTGCATTAATGATGATATAA 677
Qy 539 CAGCTGAAGGAGTCTCTGCAATTTATGTTGGATATCTACCATGGCGGTTATTTGGTAATA 718
Db 798 CTAATTAATCCGGAACCTCTGTTTTCATCGGTGGGCAATTTCCACAGGATCATAGGTAATG 857
Qy 719 AATTTGAGAAGTACTCTGTAATACTCTCTGATGATGTTTAAATGATCTCTCAAAACAAT 778
Db 858 AGTTAGAGATATTCCTGCAATGATACCTAGTAACTCAACTCAATAAGTGGACCAAT 917
Qy 779 CTGCTTCAGTACTCTTCAGCTTGGATCTTTGGCGGAGAAATTTGGAATGAGGTTCACT 838
Db 918 TTGCAACAGTAACTAAATGTTGTCTACTTTGGTTTGAACCTTCGAGGAAGATTTAACT 977
```

## RESULT 12

US-10-062-624-1

; Sequence 1, Application US/10062624

; Patent No. US20020115840A1

; GENERAL INFORMATION:

; APPLICANT: Walker, David H.

; APPLICANT: McBride, Jere W.

; APPLICANT: Yu, Xue-Jie

; TITLE OF INVENTION: Homologous 28-Kilodalton Immunodominant Protein

; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof

; FILE REFERENCE: D6152CIP2/D1

; CURRENT APPLICATION NUMBER: US/10/062,624

; PRIORITY FILING DATE: 2002-01-31

; PRIORITY FILING DATE: 2000-09-12

; NUMBER OF SEQ ID NOS: 46

; SEQ ID NO 1

; LENGTH: 1607

```
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: nucleic acid sequence of E. canis p28-7
US-10-062-624-1
```

Query Match 28.4%; Score 238.4; DB 12; Length 1607;

Best Local Similarity 57.7%; Pred. No. 3.6e-41;

Matches 486; Conservative 0; Mismatches 346; Indels 10; Gaps 3;

```
Qy 1 ATGAATTATAAGAAAAATTTCTAGTAAGAGCGGTTAAATCTCATTAATATGCTCAATCTTACCA 60
Db 146 ATGAATTGCAAAAAATTTCTATAAACAACGCTAATATATCATTAATATGCTCAATCTTACCA 205
Qy 61 TATCAGCTCTTTTTCAGATCCCTGATAGGTTCAAGAACTAATGATATAACAAGAGGCTTCTTAC 120
Db 206 AGCATATCTTTTCTGATACATATACAAGATGGTAACATGGGTGGTAA-----CTTCTAT 259
Qy 121 ATTAGTCAAAAGTACAATCCAAGTATATCACATTTTAGAAAAATTTCTGCTGAAGA--AA 178
Db 260 ATTAGTGAAGATGATGACCAAGTGTCTCACATTTTGGTAGCTTCTCAGCTAAAGAAGAA 319
Qy 179 CTCCTATTAAATGAACAATTTCTCTCACTAAAAAAGTTTTCGGACTAAAGAAGATGGTG 238
Db 320 AGCAATCAACTGTTGGAGTTTTCGGATTAAACAATGATTTGGATGGGAAGTCCAACTAT- 378
Qy 239 ATATAACAAAAAGACGATTTTACAAGAGTAGTCCAGGCATTCATTTTCAAAATAACT 298
Db 379 -TAAGAATAACACGCTGACTTACTGTTCCAACTATTGTTTCAGATACGGAACAATC 437
Qy 299 TAATATCAGGATTTTTCAGGAAGTATTTGTTACTCTATGCTAGTATTAATGATCTCTCAAA 538
Db 438 CATTTCTAGGTTTTCAGGAGCTATCGGTTTACTCAATGGTGGCCCAAGAATAAGAAATTCG 497
Qy 359 AAGCTGCATATCAACAATTTAATCCAAAAACCCGATACCAATGATATGATATGATGGTG 418
Db 498 AATATCTTTATGAAGCATTCGACGTAAAGATGCTTAATATCAATATCAAAATGACGGC 557
Qy 419 AATACTATAAACATTTTTCGATATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTAG 478
Db 558 ACAGTACTCGCTCTATCTCATACACATCGGCAGCCATGGAAGCTGATAAATTTGTCT 617
Qy 479 TACTTAAAAATGACGGCATAACTTTTATGCTAGTATTAATGATCTCTCAAAACAAT 538
Db 618 TCTTAAAAACGAAGGTTAATTCGACATATCACTTGCATTAATGATGATATAA 677
Qy 539 CAGCTGAAGGAGTATCTTTTCGATACCATATGCTAGTATGATAGGAGTATAGGAGTCTTATCA 598
Db 678 TAAATGACAAAGTACCTGTTTCTCTTATATATGCGCAGGATTTGGTACTGATTTGATTT 737
Qy 599 CTATTTTAAAGACCTCAATCTAAANNTTGTCTTACCAAGGAAAAATAGGTATTAGTTACC 658
Db 738 CTATGTTTGAAGCTACAAGTCTTAAATTTCTTACCAAGGAAAACTGGGCATTAGTTACT 797
Qy 659 CTATCACCAGGAAGTCTCTGCAATTTATGTTGGATATCTACCATGGCGGTTATTTGGTAATA 718
Db 798 CTAATTAATCCGGAACCTCTGTTTTCATCGGTGGGCAATTTCCACAGGATCATAGGTAATG 857
Qy 719 AATTTGAGAAGTACTCTGTAATACTCTCTGATGATGTTTAAATGATCTCTCAAAACAAT 778
Db 858 AGTTAGAGATATTCCTGCAATGATACCTAGTAACTCAACTCAATAAGTGGACCAAT 917
Qy 779 CTGCTTCAGTACTCTTCAGCTTGGATCTTTGGCGGAGAAATTTGGAATGAGGTTCACT 838
Db 918 TTGCAACAGTAACTAAATGTTGTCTACTTTGGTTTGAACCTTCGAGGAAGATTTAACT 977
```

## RESULT 13

US-10-062-051-5



```
; Sequence 5, Application US/10062051
; Publication No. US20030073095A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/10/062,051
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US/09/660,587
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 5
; LENGTH: 840
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; NAME/KEY: mat_peptide
; OTHER INFORMATION: nucleic acid sequence of p28-6
US-10-062-051-5

Query Match      27.9%; Score 234; DB 9; Length 840;
Best Local Similarity 59.0%; Pred. No. 2.4e-40;
Matches 503; Conservative 0; Mismatches 325; Indels 24; Gaps 5;

Qy      1 ATGAATTAAGAAAATCTAGTAAGAGCGGTTAAATCTCATTAAATGCTCAATCTTACCA 60
Db      1 ATGAATTCGCAAAAATTTCTTATAACAACACTGCAATTAATGTCTATTAATGCTCAATCTTACCA 60

Qy      61 TATCAGTCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120
Db      61 AGCATATCTTTTCTGATACTATACAAG-----ACGATAACACTGGTAGTCTTCTAC 111

Qy      121 ATTAGTGCAGAAAGTACAAATCCAAAGTATATCACATTTTAGAAAAATCTCTGCTGGAAGA--AA 178
Db      112 ATCAGTGGAAAATATGTATCCAAAGTGTTCACATTTTGGTGTGTTTCTCAGCTAAAGAGAA 171

Qy      179 CTCCTATTAAATGGAAACAAATTTCTCTCACTAAAAAGTTTTTCGGACTAAAGAAAGATGGTG 238
Db      172 AGAAACTCAACTGTGGAGTTTTTGGATTAAAAACATGATTGGAATGGAGGTACAATATCT 231

Qy      239 ATATAACAAAAAAGACGATTTTACAGAGTAGTCTCCAGGCATTTGATTTTC--AAATTAAC 297
Db      232 AACTCTTCTCCAGAAAATATATTACAGATTCGTTTAAATATCGAATAACAAAC 291

Qy      298 TTAATATCAGGATTTTTCAGGAAGTATTGGTTACTCTATGGACGACCAAGAAATAGAACTT 357
Db      292 CCATTCTTAGGGTTTGCAGGAGCTATTGGTTATTCAATGGGTGGCCCAAGNATAGAACTT 351

Qy      358 GAAGCTGCATATCAACAATTTTAAATCCAAAAAACCAGTAACCAATGATCTGATTAATGGT 417
Db      358 GAAGCTGCATATCAACAATTTTAAATCCAAAAAACCAGTAACCAATGATCTGATTAATGGT 417

Qy      418 GAATACTATAAACAATTTTGCATTTCTCGTAAAGATGCAATG-----GAAGATCAG 468
Db      412 CACAGATCTGTCTTTATCTCATCATATAGTTTCAGCAACAGCATGCTCTCCGCAAGTAAC 471

Qy      469 CAATATGATGACTTTAAAAATGACGGCATAACTTTTATGTCATGATGGTTTAATCTTGC 528
Db      472 AAATTTGTTTCTTAAAAAATGAAGGTTAATTTGACTTATCATTTATGATAAATGTCATGC 531

Qy      529 TATGACATTACAGCTGAAGGATGATCTTTCTGATACCATATGATGTCAGGATATAGGACCA 588
Db      532 TATGACATAAATAATGAAGGAATGCTTTTTCACCTTATATTGTCAGGTTGGTACT 591

Qy      589 GATCTTATCACTATTTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGAAAAATAGGT 648
Db      592 GATGTTGTTTCCATGTTTGAAGCTATAAATCTTAAATTTCTTACCAAGGAAAACTAGGA 651
```

```
US-10-062-920-5
; Sequence 5, Application US/10062920
; Publication No. US20030096250A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/10/062,920
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US/09/660,587
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 5
; LENGTH: 840
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; NAME/KEY: mat_peptide
; OTHER INFORMATION: nucleic acid sequence of p28-6
US-10-062-920-5

Query Match      27.9%; Score 234; DB 9; Length 840;
Best Local Similarity 59.0%; Pred. No. 2.4e-40;
Matches 503; Conservative 0; Mismatches 325; Indels 24; Gaps 5;

Qy      1 ATGAATTAAGAAAATCTAGTAAGAGCGGTTAAATCTCATTAAATGCTCAATCTTACCA 60
Db      1 ATGAATTCGCAAAAATTTCTTATAACAACACTGCAATTAATGTCTATTAATGCTCAATCTTACCA 60

Qy      61 TATCAGTCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120
Db      61 AGCATATCTTTTCTGATACTATACAAG-----ACGATAACACTGGTAGTCTTCTAC 111

Qy      121 ATTAGTGCAGAAAGTACAAATCCAAAGTATATCACATTTTAGAAAAATCTCTGCTGGAAGA--AA 178
Db      112 ATCAGTGGAAAATATGTATCCAAAGTGTTCACATTTTGGTGTGTTTCTCAGCTAAAGAGAA 171

Qy      179 CTCCTATTAAATGGAAACAAATTTCTCTCACTAAAAAGTTTTTCGGACTAAAGAAAGATGGTG 238
Db      172 AGAAACTCAACTGTGGAGTTTTTGGATTAAAAACATGATTGGAATGGAGGTACAATATCT 231

Qy      239 ATATAACAAAAAAGACGATTTTACAGAGTAGTCTCCAGGCATTTGATTTTC--AAATTAAC 297
Db      232 AACTCTTCTCCAGAAAATATATTACAGATTCGTTTAAATATCGAATAACAAAC 291

Qy      298 TTAATATCAGGATTTTTCAGGAAGTATTGGTTACTCTATGGACGACCAAGAAATAGAACTT 357
Db      292 CCATTCTTAGGGTTTGCAGGAGCTATTGGTTATTCAATGGGTGGCCCAAGAAATAGAACTT 351

Qy      358 GAAGCTGCATATCAACAATTTTAAATCCAAAAAACCAGTAACCAATGATCTGATTAATGGT 417
Db      352 GAAGTCTGTACGAGACATTCGATGTGAAAATCAGAACTAATATTAAAGACGGGCA 411

Qy      418 GAATACTATAAACAATTTTGCATTTCTCGTAAAGATGCAATG-----GAAGATCAG 468
Db      412 CACAGATCTGTCTTTATCTCATCATATAGTTTCAGCAACAGCATGCTCTCCGCAAGTAAC 471

Qy      469 CAATATGATGACTTTAAAAATGACGGCATAACTTTTATGTCATGATGGTTTAATCTTGC 528
Db      472 AAATTTGTTTCTTAAAAAATGAAGGTTAATTTGACTTATCATTTATGATAAATGTCATGC 531

Qy      529 TATGACATTACAGCTGAAGGATGATCTTTCTGATACCATATGATGTCAGGATATAGGACCA 588
Db      532 TATGACATAAATAATGAAGGAATGCTTTTTCACCTTATATTGTCAGGTTGGTACT 591

Qy      589 GATCTTATCACTATTTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGAAAAATAGGT 648
Db      592 GATGTTGTTTCCATGTTTGAAGCTATAAATCTTAAATTTCTTACCAAGGAAAACTAGGA 651
```

```

Db      352 GAAGTCTCTACGAGACATTCGATGTGAATAATCAGAAATATTAATAAGACGGCGCA 411
Qy      418 GAATACTATAAACAATTTGCTATATCTCTAAAGATGCAATG-----GAAGATCAG 468
Db      412 CACAGATACGTGCTTTATCTCATCATAGTTCAGCAACAAGCATGCTCTCCGCAAGTAAC 471
Qy      469 CAATATGTAGTACTTAAATAACACGGCATAAATTTATGCTCATTTGATGTTAATCTTGC 528
Db      472 AAATTTGTTTCTTAAATAAATGAAGGTTAAATGACCTTATCATCAATTTATGATAAATCATG 531
Qy      529 TATGACATTACAGCTGAAGGATGATCTTTTCGTACCATATGCTATGTCAGGATATAGAGCA 588
Db      532 TATGACATAAATTAATGAAGGATGCTTTTTCACCTTATATTTGTCAGGCTTGGTACT 591
Qy      589 GATCTTATCATATTTTAAAGACCTCAATCTAAATAATTTGCTTAAAGGAAAAATAGGT 648
Db      592 GATGTTGTTCCATGTTGAAGCTATAAATCTTAAATTTCTTAAAGGAAAAATCTAGGA 651
Qy      649 ATTAGTTACCTTATCACACAGAGTCTCTGCAATTTATGTTGTTGATACCTACCATGCGTT 708
Db      652 TTAGGTTATAGTATAAGTTTGAAGGCTCTGTTTATCGGTGGACACTTTTCACAGAGTC 711
Qy      709 ATTGTTAATAAATTTGAGAAGATACCTGTAATAAATCTCTGTAGTATTAATGATGCTCT 768
Db      712 ATAGTATGAATTTAGACATCCCTGCTATGTTCTTGTAGTGAATCAA--ATCTTCCA 768
Qy      769 CAAACACATCTGCTTCAGTAACTCTTGACGTTGGATACCTTTGGCGGAGAAATTTGGAATG 828
Db      769 GAAACCAATTTGCAATAGTAACACTAAATGTGTGTCACCTTTGGCATAGAACTTTGGAGGA 828
Qy      829 AGGTTTCACTTC 840
Db      829 AGATTTAACTTC 840

```

```

RESULT 15
US-10-062-624-5
; Sequence 5, Application US/10062624
; Patent No. US20020115840A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-Kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2/D1
; CURRENT FILING DATE: 2002-01-31
; PRIOR FILING DATE: 2000-09-12
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 5
; LENGTH: 840
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; NAME/KEY: mat_peptide
; OTHER INFORMATION: nucleic acid sequence of p28-6
US-10-062-624-5

Query Match      27.9%; Score 234; DB 12; Length 840;
Best Local Similarity 59.0%; Pred. No. 2.4e-40;
Matches 503; Conservative 0; Mismatches 325; Indels 24; Gaps 5;

Qy      1 ATGAATTTATAAGAAAATTTCTAGTAAGAGCGGTTAATCTCTAATGTCATCTTACCA 60
Db      1 ATGAATTTGCAAAAATTTCTTATACAACTGCTAATGTCATCTAATGTCATCTTACCA 60
Qy      61 TATCAGTCTTTTGGAGATCTCTGATGTTCAAGAACTAATGATGATAACAAAGAGCTTCTAC 120
Db      61 AGCATATCTTTTCTGATACATATACAAG-----ACGATAACACTGGTACTTCTAC 111

```

```

Qy      121 ATTAGTCAAAGTACAATCCAAAGTATATCATCACCATTAGAAAATTTCTCTGCTGAAGA--AA 178
Db      112 ATCAGTGGAAAATATATGTAACAAGTGTTCACATTTTGGTGTCTTCTCAGCTAAGAAGAA 171
Qy      179 CTCTATTATTAATGGAACAAATTTCTCTCACTAAAAAGTTTTCGAGACTAAGAAAAGATGGTG 238
Db      172 AGAAACTCAACTGTTGGAGTTTTGGATTAAACATGATTGGAATGGAGGTACAATATCT 231
Qy      239 ATATAACAAAATAAGACGATTTTACAGAGTAGCTCCAGGCATTTGATTTTC--AAATAAC 297
Db      232 AACTCTTCTCCAGAAAATATATTCACAGTTTCAAAAATTTATTCGTTTAAATACGAAAACAA 291
Qy      298 TTAATATCAGGATTTTTCAGGAAGTATTGGTTACTCTATGGACGCAACCAAGAAATAGAATCT 357
Db      292 CCATTTTAGGTTTTCAGGAGCTATTGGTTATTTCAATGGGTGGCCCNAGAAATAGAACTT 351
Qy      358 GAAGCTGCATATCAACAATTTTAAATCCAAAAAACCCGATAACAATGATATCTGATAATGGT 417
Db      352 GAAGTTCTGTACGAGACATTCGATGTGAAAAAATCAGAAACAATAATATAAGAACCGCGCA 411
Qy      418 GAATACTATAACAATTTTGCATTTATCTCGTAAGATGCAATG-----GAAGATCAG 468
Db      412 CACAGATACCTGCTGCTTTATCTCATCATAGTTTCAGCAACAAGCATGTCTCCGCAAGTAAC 471
Qy      469 CAATATGTAGTACTTAAATAATGACGGCATAAATTTTATGTCATTTGATGTTTAAATCTTGC 528
Db      472 AAATTTGTTTCTTAAATAATGAAAGGTTAAATGACTTATCATTTATGATAAATGCAATGC 531
Qy      529 TATGACATTAACAGCTGAAGGATGATCTTTTCGTACCATATGATGTCAGGATATAGGAGCA 588
Db      532 TATGACATAAATAATGAAGGAATGCTTTTTCACCTTATATTTGTCAGGTTGTTGGTACT 591
Qy      589 GATCTTATCATATTTTAAAGACCTCAATCTAAATAATTTGCTTACCAAGGAAAAATAGGT 648
Db      592 GATGTTGTTCCATGTTTGAAGCTATAAATCCTAAAAATTTCTTACCAAGGAAAAATAGGA 651
Qy      649 ATTAGTTACCTTATCACACAGAGTCTCTGCAATTTTATGTTGATGATACCTACCTGCGCTT 708
Db      652 TTAGTTATAGTATAAGTTTCAAGAGCTCTGTTTATCGGTGGACACTTTTCACAGATC 711
Qy      709 ATTGTTAATAAATTTGAGAAGATACCTGTAATAAATCTCTGTAGTATTAATGATGCTCT 768
Db      712 ATAGTATGAATTTAGAGACATCCCTGCTATGTTCTCTAGTGGATCAA--ATCTTCCA 768
Qy      769 CAAACACATCTGCTTCAGTAACTCTTGACGTTGGATACCTTTGGCGGAGAAATTTGGAATG 828
Db      769 GAAACCAATTTGCAATAGTAACACTAAATGTGTGTCACCTTTGGCATAGAACTTTGGAGGA 828
Qy      829 AGGTTTCACTTC 840
Db      829 AGATTTAACTTC 840

RESULT 16
US-10-314-639-37
; Sequence 37, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20030103991A1
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/314,701
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 37
; LENGTH: 843
; TYPE: DNA
; ORGANISM: Ehrlichia canis

```

```
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(843)
US-10-314-639-37

Query Match      27.7%; Score 232.4; DB 9; Length 843;
Best Local Similarity 58.9%; Pred. No. 5.3e-40;
Matches 502; Conservative 0; Mismatches 326; Indels 24; Gaps 5;

Qy 1 ATGAATTATAAGAAAATTTCTAGTAAGAGCGGTTAAATCTCATTAATGTCAATCTTACCA 60
Db 1 ATGAATTGCAAAAAATTTCTTATAACAACCTGCATTATGTCATTATGACTACTATGCTCCA 60
Qy 61 TATCAGTCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGATAACAAGAGGCTTTCTAC 120
Db 61 AGCATATCTTTTCTGATACTATAACAAG-----ACGATAACACCTGGTAGCTTCTAC 111
Qy 121 ATTAGTCAAAAGTACAATCCAAAGTATATCACTTTTAGAAAATTTCTGCTGAAGA--AA 178
Db 112 ATCAGTGAATAATATGTAACCAAGTGTTCACATTTTGGTGTTCACATTTTGGTGTTCAC 171
Qy 179 CTCCTATTATTAAGAAACAAATTTCTCACTAAAAAAGTTTTCGGACTAAAGAAAGATGGT 238
Db 172 AGAACTCACTGTTGGAGTTTGGATTAACAATGATGGAATGGAGTACAATATCT 231
Qy 239 ATATAACAAAAAGACGATTTTACAAGAGTAGCTCCAGGCAATGATTTTC-AAAAAAC-297
Db 232 AACTCTTCTCCAGAAAAATATATTCACAGTTCAAAATTTATTCGTTTAAATACGAAAAACA 291
Qy 298 TTAATATCAGATTTTTCAGGAGTATTTGGTGTCTATGACGACCAAGAAATAGACTT 357
Db 292 CCATTTCTAGGGTTTGCAGGAGCTATTGGTTATTCAATGGGTGGCCCAAGAAATAGAACTT 351
Qy 358 GAAGCTGCATATCAACAATTTAATCCAAAAACACCCGATAACAATGATCTGATAATGGT 417
Db 352 GAAGTCTGTACGAGACATTCGATGTGAAAAATCAGAACCAATATTAAGAACGGCGCA 411
Qy 418 GAATPACTATAACAATTTTGCATTTATCTCGTAAAGATGCAATG-----GAAGATCAG 468
Db 412 CACAGATCTGTGCTTTTCTCATCATAGTTTCAGAACAAACATGTCTCCGCGAAGTAAC 471
Qy 469 CAATATGTAGTACTTAAAAATGACGCAATACTTTTATGTCATTTGATGGTTTAACTTGC 528
Db 472 AAATTTGTTTCTTAAAAAATGAAAGGTTAAAGGGTTAAATGACTTATCATTTATGATAATGCA 531
Qy 529 TATGACATTAACAGCTGAAGGAGTATCTTTCTGACCATATGCTGCGAGGTATAGGAGCA 588
Db 532 TATGACATTAATTAAGGAATGCGCTTTTTCACCTTATATTGTGCGAGGTGTGGTACT 591
Qy 589 GATCTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGAAAAATAGGT 648
Db 592 GATGTTGTTTCCATGTTTGAAGCTATAAATCCCTAAAAATTTCTTACCAAGGAAAACTAGGA 651
Qy 649 ATTAGTTACCTTATCACCACAGCAAGTCTCTGCATTTTATTTGTTGATACTACCTACCTGGGTT 708
Db 652 TTAGGTTATAGTATAAGTTTCAAGAGCCTCTGTTTATTCGGTGGACACTTTTACAGAGTC 711
Qy 709 ATTGGTAATAAATTTGAGAAGATACCTGTATAACTCTCTGTAGTATTAAATGATGCTCCT 768
Db 712 ATAGGTAATGAATTTAGAGACATCCCTGCTATGGTTCTTAGTGGATCA---ATCTTCCA 768
Qy 769 CAAACCAACATCTGTTTCAGTAACCTTTGACGTTGGATACCTTTTGGCGGAGAAATGGAAATG 828
Db 769 GAAAAACCAATTTGCAATAGTAACACTAAATGTGTGTCACCTTTGTTTGAAGAACTTTGGAGGA 828
Qy 829 AGGTTCACTTC 840
Db 829 AGATTTAACTTC 840

Query Match      27.7%; Score 232.4; DB 12; Length 843;
Best Local Similarity 58.9%; Pred. No. 5.3e-40;
Matches 502; Conservative 0; Mismatches 326; Indels 24; Gaps 5;

Qy 1 ATGAATTATAAGAAAATTTCTAGTAAGAGCGGTTAAATCTCATTAATGTCAATCTTACCA 60
Db 1 ATGAATTGCAAAAAATTTCTTATAACAACCTGCATTATGTCATTATGACTACTATGCTCCA 60
Qy 61 TATCAGTCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGATAACAAGAGGCTTTCTAC 120
Db 61 AGCATATCTTTTCTGATACTATAACAAG-----ACGATAACACCTGGTAGCTTCTAC 111
Qy 121 ATTAGTCAAAAGTACAATCCAAAGTATATCACTTTTAGAAAATTTCTGCTGAAGA--AA 178
Db 112 ATCAGTGAATAATATGTAACCAAGTGTTCACATTTTGGTGTTCACATTTTGGTGTTCAC 171
Qy 179 CTCCTATTATTAAGAAACAAATTTCTCACTAAAAAAGTTTTCGGACTAAAGAAAGATGGT 238
Db 172 AGAACTCACTGTTGGAGTTTGGATTAACAATGATGGAATGGAGTACAATATCT 231
Qy 239 ATATAACAAAAAGACGATTTTACAAGAGTAGCTCCAGGCAATGATTTTC-AAAAAAC-297
Db 232 AACTCTTCTCCAGAAAAATATATTCACAGTTCAAAATTTATTCGTTTAAATACGAAAAACA 291
Qy 298 TTAATATCAGATTTTTCAGGAGTATTTGGTGTCTATGACGACCAAGAAATAGACTT 357
Db 292 CCATTTCTAGGGTTTGCAGGAGCTATTGGTTATTCAATGGGTGGCCCAAGAAATAGAACTT 351
Qy 358 GAAGCTGCATATCAACAATTTAATCCAAAAACACCCGATAACAATGATCTGATAATGGT 417
Db 352 GAAGTCTGTACGAGACATTCGATGTGAAAAATCAGAACCAATATTAAGAACGGCGCA 411
Qy 418 GAATPACTATAACAATTTTGCATTTATCTCGTAAAGATGCAATG-----GAAGATCAG 468
Db 412 CACAGATCTGTGCTTTTCTCATCATAGTTTCAGCAACAAACATGTCTCCGCGAAGTAAC 471
Qy 469 CAATATGTAGTACTTAAAAATGACGCAATACTTTTATGTCATTTGATGGTTTAACTTGC 528
Db 472 AAATTTGTTTCTTAAAAAATGAAAGGTTAAATGACTTATCATTTATGATAATGCAATGC 531
Qy 529 TATGACATTAACAGCTGAAGGAGTATCTTTCTGACCATATGCTGCGAGGTATAGGAGCA 588
Db 532 TATGACATTAATTAAGGAATGCGCTTTTTCACCTTATATTGTGCGAGGTGTGGTACT 591
Qy 589 GATCTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGAAAAATAGGT 648
Db 592 GATGTTGTTTCCATGTTTGAAGCTATAAATCCCTAAAAATTTCTTACCAAGGAAAACTAGGA 651
Qy 649 ATTAGTTACCTTATCACCACAGCAAGTCTCTGCATTTTATTTGTTGATACTACCTACCTGGGTT 708
Db 652 TTAGGTTATAGTATAAGTTTCAAGAGCCTCTGTTTATTCGGTGGACACTTTTACAGAGTC 711
Qy 709 ATTGGTAATAAATTTGAGAAGATACCTGTATAACTCTCTGTAGTATTAAATGATGCTCCT 768
Db 712 ATAGGTAATGAATTTAGAGACATCCCTGCTATGGTTCTTAGTGGATCA---ATCTTCCA 768
Qy 769 CAAACCAACATCTGTTTCAGTAACCTTTGACGTTGGATACCTTTTGGCGGAGAAATGGAAATG 828
Db 769 GAAAAACCAATTTGCAATAGTAACACTAAATGTGTGTCACCTTTGTTTGAAGAACTTTGGAGGA 828
Qy 829 AGGTTCACTTC 840
Db 829 AGATTTAACTTC 840

RESULT 17
US-10-059-964-37
; Sequence 37, Application US/10059964
```

QY 709 ATTGTAATAAATTTGAGAGATACCTGTATAAATCTCTGTAGTATTAAATGATGCTCT 768  
DB 712 ATAGTAATAAATTTAGAGACATCCCTGCTATGGTTCCTAGTGGATCAA---ATCTTCCA 768  
QY 769 CAACACACATCTGCTTCACTACTCTTGACGTTGATATCTTTGGCGGAGAAATTCGAATG 828  
DB 769 GAAACCAATTTGCAATAGTAACACTAAATGTTGTCACCTTTGGTTTAGAACTTGGAGGA 828  
QY 829 AGGTTCACTTC 840  
DB 829 AGATTAACCTC 840

RESULT 18  
US-10-314-639-1  
; Sequence 1, Application US/10314639  
; Publication No. US20030103991A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; FILE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/314,639  
; CURRENT FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US/09/314,701  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 846  
; TYPE: DNA  
; ORGANISM: Ehrlichia chaffeensis  
; NAME/KEY: CDS  
; LOCATION: (1)...(846)  
US-10-314-639-1

Query Match 27.7%; Score 232.4; DB 9; Length 846;  
Best Local Similarity 58.1%; Pred. No. 5.4e-40;  
Matches 495; Conservative 0; Mismatches 336; Indels 21; Gaps 4;

QY 1 ATGATTAATAAGAAATTTCTAGTAAGAGCGGTAAATCTCAATTAATGCTCACTTACCA 60  
DB 1 ATGATTAATAAGAAATTTCTAGTAAGAGCGGTAAATCTCAATTAATGCTCACTTACCT 60  
QY 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAGAACTAAATGATAAAGAGAGGCTTTTAC 120  
DB 61 GGAGTATCATTTTCGACCCAGCAGGTAGTGTATTAACGGTAAT-----TTCTAC 111  
QY 121 ATTAGTCAAGATACAAATCCAGTATATCACTTTAGAAATTTCTGCT---GAAGAA 177  
DB 112 ATCAGTGGAAATATACATCCCAAGTCTTGGAGTATTTGGAGTATTTCTGCTAAGGAAGA 171  
QY 178 ACTCCTATTAAATGGAACAAATTTCTCACTAAAGATTTTCGGACTAAAGAAAGTGT 237  
DB 172 AGAAATACAAAGTGTGGAGTGTGGAGTGAAGCAAAATTTGGAGCGGAGCGCAATATCC 231  
QY 238 GATATAACAAAAAGACGATTTTACAAGATAGTCTCCAGGCAATGATTTTCAAAATAC 297  
DB 232 AACTCTCCCAACGATGATTTCTACTGTCTCAATATTATTTCAATATGAAACAC 291  
QY 298 TTAATATCAGGATTTTTCAGAGATATTTGTTACTCTATGCGGACCAAGATAGAACTT 357  
DB 292 CGGTTTAAAGTGTTCAGGAGCTATTTGTTACTCAATGGATGTTCCAAAGATAGAGCTT 351  
QY 358 GAAGTGCATATCAAAATTTTAAATCCAAACACCGATAAATGATATGATATGATGCT 417  
DB 352 GAAGTATCTTATGAACATTTGATGTAAATAATCAAGGTAAATATTAAGATGAAGCA 411  
QY 418 GAATACATAAAACATTTTGCATTTATCTCGTAAAGATGCAATG-----GAAGATCAGCAA 471

DB 412 CATAGATATTGCTCTATCCATAACTCAGCAGCAGCATGAGTAGTGCAAGTAATAAT 471  
QY 472 TATGTAGTACTTTAAATAGCGGATTAACATTTTATGTCATTTGATGTTTAACTTCTGCTAT 531  
DB 472 TTTGCTCTTTTAAATAGGAGTATTTGACATATCATTTATGCTGAACGCGATGCTAT 531  
QY 532 GACATTACAGCTGAAGAGATCTTTTCTGATCCATATGATGCGAGGTATAGGAGCAGAT 591  
DB 532 GACGTAGTAGCGGAGGATACCTTTTCTCTTATATATGCGGAGGTATCGGTACTGAT 591  
QY 592 CTTATCACTATTTTAAAGACCTCAATTAATAATTTGCTTACCAAGGAAATAGGTATT 651  
DB 592 TTAGTATCCATGTTTGAAGCTACAAATCTTAAATTTCTTACCAAGGAAATAGGTATT 651  
QY 652 AGTTACCTATCACACAGAGTCTCTGCATTTTATGTTGATGATACCTACCTGGGCTTATT 711  
DB 652 AGCTACTCTATTAAGCCAGAGCTTCTGTGTTTATGTTGCGGCACTTTTCATTAAGGTATTA 711  
QY 712 GGTAAATAATTTGAGAAGATAC---CTGTAATAACTCTCTGTAGTATTAAATGATGCTCCT 768  
DB 712 GGGAAAGATTTAGAGATATTTCTTACTATATATCTACTGATCAACACTTGCAGGAAA 771  
QY 769 CAAACCACTCTGCTTCACTTAATCTTGCAGCTTGCAGTATGCTTGGCGGAGAAATGGAAATG 828  
DB 772 GGAACCTACCTGCAATAGTAAATCTGATGATGATGATGATGATGATGATGATGATGATG 831  
QY 829 AGGTTCACTTC 840  
DB 832 AGGTTGCTTTC 843

RESULT 19  
US-10-059-964-1  
; Sequence 1, Application US/10059964  
; Patent No. US20020120115A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20020120115A1  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; FILE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/059,964  
; CURRENT FILING DATE: 2002-01-28  
; EARLIER APPLICATION NUMBER: 09/314,701  
; EARLIER FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 846  
; TYPE: DNA  
; ORGANISM: Ehrlichia chaffeensis  
; NAME/KEY: CDS  
; LOCATION: (1)...(846)  
US-10-059-964-1

Query Match 27.7%; Score 232.4; DB 12; Length 846;  
Best Local Similarity 58.1%; Pred. No. 5.4e-40;  
Matches 495; Conservative 0; Mismatches 336; Indels 21; Gaps 4;

QY 1 ATGATTAATAAGAAATTTCTAGTAAGAGCGGTAAATCTCAATTAATGCTCACTTACCA 60  
DB 1 ATGATTAATAAGAAATTTCTAGTAAGAGCGGTAAATCTCAATTAATGCTCACTTACCT 60  
QY 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAGAACTAAATGATAAAGAGAGGCTTTTAC 120  
DB 61 GGAGTATCATTTTCGACCCAGCAGGTAGTGTATTAACGGTAAT-----TTCTAC 111  
QY 121 ATTAGTCAAGATACAAATCCAGTATATCACTTTAGAAATTTCTGCT---GAAGAA 177  
DB 112 ATCAGTGGAAATATACATCCCAAGTCTTGGAGTATTTGGAGTATTTCTGCTAAGGAAGA 171  
QY 178 ACTCCTATTAAATGGAACAAATTTCTCACTAAAGATTTTCGGACTAAAGAAAGTGT 237

Db 172 AGAATAACAAGTGGAGTGTGGAGTGAAGCAAAATCGGACGGAAGCGCAATATCC 231  
Qy 238 GATATAACAAAAAGACGATTTTACAAAGAGTAGCTCCAGGCAATGATTTTCAAAAATAAC 297  
Db 232 AACTCCTCCCAACGATGATCTCAATTTATTCATTTAATATGAACAAC 291  
Qy 298 TTAATATCAGATTTTCAAGAGTATGTTTACTCTATGAGCGGACCAAGAAATAGAACTT 357  
Db 292 CCGTTTTAGGTTTTGACAGGAGTATGTTTACTCAATGGATGGTCCCAAGAAATAGAGCTT 351  
Qy 358 GAAGTCGATATCAACATTTAAATCCAAAAACACCGATACAACTGATGATATGAT 417  
Db 352 GAAGTATCTTATGAACAATTTGATGTAATAAATCAAGTAACTATTAAGAAATGAAGCA 411  
Qy 418 GAATATAAACAATTTTGCATTTCTCGTAAAGATGCAATG-----GAAGATCAGCAA 471  
Db 412 CATAGATATTGCTCTATCCATAAATCAGCAGCAGACATGAGTAGTGCAGTAATAT 471  
Qy 472 TATGTAGTACTTAAAAATGACGGCAATACTTTTATGTCATTTGATGTTTAAATATCTTGCTAT 531  
Db 472 TTTGTCTTTCTTAAAAAATGAAGGATTAATCTGACATATCATTTATGCTGAACGCAATGCTAT 531  
Qy 532 GACATTACAGCTGAGGAGTATCTTTGTCATCATATGATGTCAGGTATAGGAGCAGAT 591  
Db 532 GACGTAGTAGGAGGACATACCTTTTCTCTTATATATGCGCAGGTATCGTACTGAT 591  
Qy 592 CTATCATCTATTTTAAAGACTCAATCTAAATTTGCTTAAAGGAAAAATAGGTATT 651  
Db 592 TTAGTATCATGTTTGAAGCTACAACTCTAAATTTCTTACCAAGGAAATGTTAGGTTA 651  
Qy 652 AGTTACCTATCACACCAAGAGTCTCTGCAATTTATGTTGGATACTACCATGGCGTTATT 711  
Db 652 AGCTACTCTATAAGCCCAAGAGCTCTGTGTTTATTTGGTGGCACTTTTCATAAGGTATTA 711  
Qy 712 GGTAAATATTTGAGAGATAC---CTGTATAACTCTGTAGTATTAATATGATGCTCCT 768  
Db 712 GGGAAAGAAATTTAGAGATATTTCTACTATAATACCTACTGGATCAACACTTTGCAAGAAA 771  
Qy 769 CAACACCATCTGCTTCAAGTAACTCTTGACGTTGGATCTTTGGCGGAGAAATTTGGAATG 828  
Db 772 GGAACACTCCCTGCAATAGTATATCTGATATATACCTACTGGATCAACACTTTGCAAGAAA 831  
Qy 829 AGGTTACCTTC 840  
Db 832 AGGTTGCTTC 843

## RESULT 20

US-10-062-994-11  
; Sequence 11, Application US/10062994  
; Patent No. US20010001661A1  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Bowie, Michael V.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Mahan, Suman M.  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Moreland, Annie L.  
; APPLICANT: Simbi, Bigboy H.  
; APPLICANT: Whitmore, William M.  
; APPLICANT: Alleman, Arthur R.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and  
; FILE REFERENCE: UP-167XC3  
; CURRENT APPLICATION NUMBER: US/10/062,994  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: 09/533,662  
; PRIOR FILING DATE: 2000-04-21  
; PRIOR APPLICATION NUMBER: 09/337,827  
; PRIOR FILING DATE: 1999-06-22  
; PRIOR APPLICATION NUMBER: 08/953,326

; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 08/733,230  
; PRIOR FILING DATE: 1996-10-17  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 830  
; TYPE: DNA  
; ORGANISM: Ehrlichia chaffeensis  
US-10-062-994-11

Query Match 27.2%; Score 228.6; DB 12; Length 830;  
Best Local Similarity 58.2%; Pred. No. 3.4e-39;  
Matches 487; Conservative 0; Mismatches 329; Indels 21; Gaps 4;  
Qy 1 ATGAATTTATAAGAAAAATTTCTAGTAAGAAAGCGGTTAATCTCATTAATGTCTAATCTTACCA 60  
Db 3 ATGAATTTACAANAAGTTTTTCATNAACAAGTGCAATGATATCATTAATATCTTCTCTACCT 62  
Qy 61 TATCAGTCTTTTTCAGATCCTGTAGGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120  
Db 63 GGAGTATCATTTTCCGACCCAGCAGGTAGTGGTATTAAACGGTAAT-----TTCTAC 113  
Qy 121 ATTAGTCAAAAGTACAAATCCAAGTATATCACATTTTAGAAAAATTTCTCTGCT---GAAGAA 177  
Db 114 ATCAGTGGAAAAATACATGCCAAGTGTCTCGCATTTTGGAGTATTTCTCTGCTAAGGAAGAA 173  
Qy 178 ACTCCTATTAAATGAAGCAAAATTTCTCTCACTAAAAAAGTTTTTCGACCTAAAGAAAGATGGT 237  
Db 174 AGAATACAACAGTTGGAGTGTTTGGACTGAAGCAAAATTTGGGACGGAGCGCAATATCC 233  
Qy 238 GATATAACAANAAGACGATTTTACAAAGAGTAGCTCCAGGCATTTGATTTTCAAAAATAAC 297  
Db 234 AACTCCTCCCAACGATGATTTCACTGTCTCAATATTTCATTTAAATATGAACAAC 293  
Qy 298 TTAATATCAGATTTTTCAGAAAGTATTTGGTGTACTCTATGACGCAACCAAGAAATAGAACTT 357  
Db 294 CCGTTTTTAGGTTTTGCGAGGAGCTATTTGGTGTACTCAATGGATGGTCCCAAGAAATAGAGCTT 353  
Qy 358 GAAGTCGATATCAACAAATTTTAAATCCAAAAACACCGATTAACATGATCTGATAATGGT 417  
Db 354 GAAGTATCTTATGAACAATTTGATGTAANAATCAAGTAACTATTAAGAAATGAAGCA 413  
Qy 418 GAATATAAACAATTTTGCATTTATCTCGTAAAGATGCAATG-----GAAGATCAGCAA 471  
Db 414 CATAGATATTGCTCTATCCATAACTCAGCAGCAGACATGAGTAGTGCAGTAATAT 473  
Qy 472 TATGTAGTACTTAAAAATGAGGCAATACTTTTATGTCATTTGATGTTGTTAATATCTTCTAT 531  
Db 474 TTTGTCTTTCTTAAAAAATGAAGGATTAATTTGACATATCATTTATGCTGAACGCAATGCTAT 533  
Qy 532 GACATTACAGCTGAAGGAGTATCTTTGTCATCATATGTCATGCGAGGTATAGGAGCAGAT 591  
Db 534 GACGTAGTAGGAGGACATACCTTTTCTCTTATATATGCGCAGGTATCGGTACTGAT 593  
Qy 592 CTATCATCTATTTTAAAGACCTCAATCTAAAAATTTGCTTTACCAAGGAAAAATAGGTATT 651  
Db 594 TTAGTATCCATGTTTGAAGCTACAAATCTTAAATTTCTTACCAAGGAAATGTTAGGTTA 653  
Qy 652 AGTTACCTATACACCAAGGATCTCTGCAATTTTATGTTGGTGGATACTACCATGGCGTTATT 711  
Db 654 AGCTACTCTATAAGCCCAAGGATTTCTGTGTTTATTTGGTGGCACTTTTCATAAGGTAATA 713  
Qy 712 GGTAAATATTTGAGAGATAC---CTGTATAACTCTCTGTAGTATTAATATGATGCTCCT 768  
Db 714 GGGAAACGAATTTAGAGATATTTCTTACTATTAATACCTACTGGATCAACACTTTGCAAGAAA 773  
Qy 769 CAACACCATCTGCTTCAAGTAACTCTTGAGCTTTGGATACTTTTGGCGGAGAAATTTGGA 825  
Db 774 GGAACACTACCTGCAATAGTAACTAGTATGTCGATGTATGCCACTTTTGGAAATAGAAATGGA 830

RESULT 21

US-10-062-994-11  
; Sequence 11, Application US/10062994  
; Patent No. US20020132789A1  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Bowie, Michael V.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Mahan, Suman M.  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Moreland, Annie L.  
; APPLICANT: Simbi, Bigboy H.  
; APPLICANT: Whitmire, William M.  
; APPLICANT: Allemen, Arthur R.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and  
; TITLE OF INVENTION: Methods of use  
; FILE REFERENCE: UP-167XC3  
; CURRENT APPLICATION NUMBER: US/10/062,994  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: 09/533,662  
; PRIOR FILING DATE: 2000-04-21  
; PRIOR APPLICATION NUMBER: 09/337,827  
; PRIOR FILING DATE: 1999-06-22  
; PRIOR APPLICATION NUMBER: 08/953,326  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 08/733,230  
; PRIOR FILING DATE: 1996-10-17  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 830  
; TYPE: DNA  
; ORGANISM: Ehrlichia chaffeensis  
US-10-062-994-11

Query Match 27.2%; Score 228.6; DB 12; Length 830;  
Best Local Similarity 58.2%; Pred. No. 3.4e-39;  
Matches 487; Conservative 0; Mismatches 329; Indels 21; Gaps 4;

QY 1 ATCAATATTAAGAAATTTCTAGTAAGAGCGGTTAATCTCAATTAATGTCATCTTACCA 60  
DB 3 ATGAATTTACAAAAGGTTTTCATAACCAAGTCGATTGATATCAATTAATATCTTCTCTACT 62

QY 61 TATCAGTCTTTTCAGATCTCTGAGTGTCAAGAACTAATGATAACAAGAGGCTTTTAC 120  
DB 63 GGAGTATCTTTTCCAGCCAGCGAGTAGTGGTATTAACGGTAAT-----TTCTAC 113

QY 121 ATTAGTGCAAGTATCAATCAAGTATATCACACTTTAGAAAATTTCTTGCT---GAAGAA 177  
DB 114 ATCAGTGGAAAATACATGCCAGTGTCTCGCATTTTGGAGTATTTCTGCTAAGGAAGAA 173

QY 178 ACTCCTATTATGGAACAATTTCTCACTAAAAAAGTTTTCGGACTAAAGAAAGATGGT 237  
DB 174 AGAATATCAACAGTGTGGAGTGTGTGGACTGAAGCAAAATTTGGGACGCGCAATATCC 233

QY 238 GATATAACAAAAGAGAGATTTTACAAGTAGTCTCCAGGCAATGATTTTCAAAATAAC 297  
DB 234 AACTCTCCCAACAGATGATTCATCTGCTCAATTAATTCATTAATATGAAAACAC 293

QY 298 TTAATATCAGGATTTTTCAGGAAGTATTTGTTACTCTATCGGACGCGCAAGAAATAGACTT 357  
DB 294 CCGTTTTTAGTGTTCAGGAGCTATTGTTACTCAATCGATGGTCCAAAGATAGAGCTT 353

QY 358 GAAGCTGCATATCAACAATTTAATCAAAAACACCGATAACAATGATGATCTGATATGGT 417  
DB 354 GAAGTATCTTATGAACAATTTGATGTAATAAAATCAAGGTAAACAATTTAAGAATGAAGCA 413

QY 418 GAATCTATATAACATTTTTCGATTATCTCGTAAAGATGCAATG-----GAAGATCAGCAA 471  
DB 414 CATAGATATTTGCTCTATATCCCACTACTCAGCAGCAGATGATGTCAGAGTAAAT 473

QY 472 TATGTAGTACTTAAAAATGACGCATACATTTTATGTCATTTGATGGTAAATCTTGCTAT 531

```
Db 295 TTTAAATATGAACAAATCCATTTTGGGTTTGCAGGAGTATTGGCTACTCAATGGGT 354
Qy 340 GGACCAAGATAGAACTTTGAAGCTGCATATCAACAAATTTAAATCCAAAAAACCACCGATAAC 399
Db 355 GGTCCAAGGGTAGAGTTTGAAGTGTCTTACGAAATATTTGATGTAAAAAACCAGGTAAAC 414
Qy 400 AATGATCTGATATGATGGTGAATCTATATAAACAATTTTGGCAATTTCTCGTAAAGATGCAATG 459
Db 415 AGTTCAAGAAGACGCTGCACAAATATTGGCGTTTATCAAGACACACACCGGAGGTATGCCA 474
Qy 460 GAAG-----ATCAGCAATATCTAGTACTTAAATATGAAGATGCGCATTAATGTTTATGTC 510
Db 475 CAAGCCGCTCATCAAAATAAATTTGTTCTTCTTAAAAAATGAAGGATTAATTTGACATATCA 534
Qy 511 TTGATGTTAAATCTGCTATGCAATATACAGCTGAAGGAGTATCTTTCGTACCATATGCA 570
Db 535 CTTATGATAAAGCGATGTTATGATATACAAATCGACAGCATGCCAATTTTCCATATATA 594
Qy 571 TGTGAGGTATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCT 630
Db 595 TGTGAGGTATGTTAGTGTAGTCTTGTGATTTTGAAGCTTACAAATCTTAAAAATTTCT 654
Qy 631 TACCAAGCAAAATAGGTATTAGTTTACCTATCACACGAGTCTCTGCAATTTATTTGGT 690
Db 655 TATCAAGGAAATATAGGTATGATTTACTTCCATAAGCCCAAGGACATCTGTTTGTGGGA 714
Qy 691 GGATACTACCATGCGTTTATGGTAAATAAATTTGAGAAGATACCTGTAAATCTCCTGTA 750
Db 715 GGACACTTTCACAGATTATAGGTAAATGAATTTAAAGACATCTCTGCAATTAATCTCTGCT 774
Qy 751 GTATTAAATGATGCTCTCAACACATCTGCTTCAAGTAACTCTTTGAGTTGGATCTTT 810
Db 775 GGAGCAACAGAAATTAAGGACACACAGTTTACACAGTAAACATTAAACATATGCCACTTC 834
Qy 811 GGCGGAGAAATTTGAATCAGGTTACCTT 839
Db 835 GGACTAGAGCTTGGAGGCGGTTTACTTTT 863
```

## RESULT 23

```
US-10-059-964-31
; Sequence 31, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20020120115A1io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; EARLIER FILING DATE: 2002-01-28
; EARLIER APPLICATION NUMBER: 09/314,701
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(867)
US-10-059-964-31
```

```
Query Match: 25.9%; Score 217.4; DB 12; Length 867;
Best Local Similarity 57.8%; Pred. No. 8.2e-37;
Matches 502; Conservative 0; Mismatches 331; Indels 36; Gaps 5;

Qy 1 ATGAATTATAAGAAAATCTAGTAGAAGCGCGTTAATCTCATTAATGTCATCTTACCA 60
Db 1 ATGAATTGCAAAAGATTTTTCATAGCAAGTGCATGATATCACTAATGTCCTTCTTACCT 60
```

```
Qy 61 TATCAGTCTTTTGGCAGATCTCTGTAGGTTCAAGAACTAATATGATAACAAAGAGGCTTCTTAC 120
Db 61 AGCGTATCTTTTCTGAATCAATACATGAAG-----ATAATATAAATGGTAACTTTTAC 114
Qy 121 ATTAGTGCAAAGTACAAATCCAAAGTATATCACACTTTAGAAAAATTTCTCTGCTGAAGAACT 180
Db 115 ATTAGTGCAAAGTATATGCCAAGTGCCTTCACACTTTGGCGTATTATTCTCAATTTAAAGAGAG 174
Qy 181 CCTATTATGAAC- -AAATTTCTCTCACTAAAAAGTTTTCGGAC-----TAAAG 228
Db 175 AAAAACCAACAACTGGAGTTTTCGGATTAAAAACAAGTTGGGACGGAGCAACATAAAG 234
Qy 229 AAAGATGGTATATAACAAAAAAGAGCAATTTTCAAGAGTAGTCTCCAGGCAATGATTT- 287
Db 235 GATCAAGCAGCAGCCACACAAATAGACCAAGTACAATATTCTCCATTTCAATTTAATTC 294
Qy 288 -----TCAAAATAAATTAATATCAGGATTTTTCAGGAGTATTTGGTACTCTATGAC 339
Db 295 TTTAAATATGAAAAACAATCCATTTTATAGGGTTTTCAGGAGCTATTGGCTACTCAATGGGT 354
Qy 340 GGACCAAGAAATAGAACTTTGAAGCTGCATATCAACAAATTTAATCCAAAAAACCACGATAAC 399
Db 355 GGTCCAAGGGTAGAGTTTGAAGTGTCTTACGAAATATTTGATGTAAAAAACCAGGTAAAC 414
Qy 400 AATGATCTGATATGATGGTGAATACTATAAACAATTTTGCATTTCTCGTAAAGATGCAATG 459
Db 415 AGTTCAAGAAGCAGTCTCAAAATAATTGGCTTTATCAAGACACACACCGGAGGTATGCCA 474
Qy 460 GAAG-----ATCAGCAATATGATAGTACTTAAATATGACGCAATTAATTTATGTC 510
Db 475 CAAGCCGCTCATCAAAATAAATTTGTTCTTCTTAAAAAATGAAGGATTAATTTGACATATCA 534
Qy 511 TTGATGTTAATATCTGCTATGCAATTACAGCTGAAGGAGTATCTTTCGTACCATATGCA 570
Db 535 CTTATGATAAAGCGATGTTATGATATACAAATCGACAGCATGCCAATTTTCTCCATATATA 594
Qy 571 TGTGAGGTATGAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCT 630
Db 595 TGTGAGGTATGTTAGTGTAGTCTTGTCTTCTTAAAAAATGAAGGATTAATTTTGTGGGA 654
Qy 631 TACCAAGGAAATATAGTATTAGTTTACCTATCACACGAGTCTCTGCAATTTATTTGGT 690
Db 655 TATCAAGGAAATATAGGTATGATTTACTTCCATAAGCCCAAGGACATCTGTTTGTGGGA 714
Qy 691 GGATACTACCATGCGCTTATTGGTAAATAAATTTGAGAAGATACCTGTAAATCTCCTGTA 750
Db 715 GGACACTTTCACAGATTATAGGTAAATGAATTTAAAGACATCTCTGCAATTAATCTCTGCT 774
Qy 751 GTATTAAATGATGCTCTCAACACATCTGCTTCAAGTAACTCTTTGAGTGGATCTTT 810
Db 775 GGAGCAACAGAAATTAAGGACACACAGTTTACACAGTAAACATTAAACATATGCCACTTC 834
Qy 811 GGCGGAGAAATTTGAATCAGGTTACCTT 839
Db 835 GGACTAGAGCTTGGAGGCGGTTTACTTTT 863
```

## RESULT 24

```
US-10-314-639-7
; Sequence 7, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20030103991A1io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/314,701
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 7
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(861)
US-10-314-639-7

Query Match      25.3%; Score 212.8; DB 9; Length 861;
Best Local Similarity 57.3%; Pred. No. 7.8e-36;
Matches 492; Conservative 0; Mismatches 342; Indels 24; Gaps 5;

QY 1 ATGAATTATAGAAAATCTAGTAGAGACGGCTTAATCTCAATTAATGTCATCTTACCA 60
DB 1 ATGAATTCGGAATAATTTTATATACAACTGCAATTAACATTAATGTCCTTCTTACCT 60
QY 61 TATCAGTCTTTTGCAGATCTCTAGGTTCAAGAACTTAATGATAACAAGAGGCTTCTAC 120
DB 61 GGAATATCACTTTCTGATCCAGTACAG-----GATGACAACTTAGTGGTAATTTCTAC 114
QY 121 ATTAGTCGAAGTACAATCAAGTATATCACACTTTAGAAAATTTCTCTGCT---GAAGAA 177
DB 115 ATCAGTGGAAAGTATATGCAAGCGCTTCGCATTTTGGAGTTTCTTGCCCAAGGAAGA 174
QY 178 ACTCCTTATTAATGGAAACAATTTCTCTCACTAAAGATTTTCGGACTAAAGAAAGATGGT 237
DB 175 AGAAATACAAACAGTTGGAGTATTTGGAAATAGACGAAGATTGGGATAGATGTGTAATATCT 234
QY 238 GATATAACAAAAAGACGATTTTACAAGAGTAGCTCCAGGCAATTTGATTTTCAAAATAAC 297
DB 235 AGAACCACTTTAAGCGATATATTCACCGTTCCAAATATTATTAAAGTATGGAATAAT 294
QY 298 TTAATATCAGGATTTTCAGGAAGTATTTGGTTACTCTATGACGCGACCAAGATAGAACTT 357
DB 295 CTATTTTCAGGATTTGCAGGAGCTATTGGCTACTCAATGGATGGCCCAAGATAGAGCTT 354
QY 358 GAAGCTGCATATCAACAAATTTTATCCAAAACACCGATAACAATGATGATGATG--- 415
DB 355 GAAGTATCTTATGAAGACATTCGATGTATAAAATCAAGGTAAACAATTAAGAACAAGCA 414
QY 416 -GTGAATACATAAACAATTTTGCATTTATCTCTGTAAGATGCAATGGAAGATCAGCAA--- 471
DB 415 CATAGATATTATGCTCTGTGCCATCTCTCGSCACAGACACAGATAGATGGTGCAGGC 474
QY 472 -----TATGTAGTACTTAAATAAGACGCGATAAATTTTATGTCATTTGATGTTTAACT 525
DB 475 AGTGGGTCTGTCTTTCTAATAAATGAAGGACTACTTGATAAATCAATTTATGCTGAACGCA 534
QY 526 TGCTATGACATTTACAGCTGAAGGAGTATCTTTTCGTACCATATGTCATGTCAGGTATAGGA 585
DB 535 TCTTATGATGTAATAAGTGAAGGACATACCTTTCTCTTATATATGTCAGGTATTGGT 594
QY 586 GCAGATCTTATCAGTATTTTAAAGACCTCAATCTTAAATTTTGGCTTACCAGGAAATAA 645
DB 595 ATTGATTTAGTATCCATGTTGAGCTATAAATCTTAAATTTCTTATCAAGGAAATA 654
QY 646 GGTATTAGTACCTATACACAGAGTCTCTGCAATTTTATTTGTTGGATPACTACCATGGC 705
DB 655 GGCTTAAGTTACCTATAGCCAGGAGCTCTCTGTGTTTATTTGGTGGACATTTTCATAAG 714
QY 706 GTTATTTGTAATAATTTGAGAGATACCTGTATAAATCTCTGTAGTATTAATGATGCT 765
DB 715 GTGATAGGAAACGAATTTTAGAGATATTTCTTACTATGATACCTAGTGAATCAGCGCTTGA 774
QY 766 CCTCA---AACCACATCTCTTTCAGTAACTCTTGAGTTGGATCTTTGGCGGAGAAAT 822
DB 775 GGAAGAAGGAATACCTCTCAATAGTAACACTGGAGCTGTTCTACTTTGGCATAGAACTT 834
QY 823 GGAATGAGTTTCACTTTC 840
DB 835 CGAGGAAGGTTTAACTTC 852
```

```
RESULT 25
US-10-062-994-8
; Sequence 8, Application US/10062994
; Patent No. US20010001661A1
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Bowie, Michael V.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Mahan, Suman M.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Moreland, Annie L.
; APPLICANT: Simbi, Bigboy H.
; APPLICANT: Whitmore, William M.
; APPLICANT: Alleman, Arthur R.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and
; FILE REFERENCE: UP-167XC3
; CURRENT APPLICATION NUMBER: US/10/062,994
; PRIORITY FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 09/533,662
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 09/337,827
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 08/953,326
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 08/733,230
; PRIOR FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
US-10-062-994-8
```

```
Query Match      25.3%; Score 212.8; DB 12; Length 861;
Best Local Similarity 57.3%; Pred. No. 7.8e-36;
Matches 492; Conservative 0; Mismatches 342; Indels 24; Gaps 5;

QY 1 ATGAATTATAGAAAATCTAGTAGAGACGGCTTAATCTCAATTAATGTCATCTTACCA 60
DB 3 ATGAACTCGGAAAATTTTATTAACAACCTGCAATTAACATTTACTTAATGCTTCTTACCT 62
QY 61 TATCAGTCTTTTGCAGATCTCTGTTAGGTTCAAGAACTTAATGATAACAAGAGGCTTCTAC 120
DB 63 GGAATATCACCTTTCTGATCCAGTACAG-----GATGACAACTTAGTGGTAATTTCTTAC 116
QY 121 ATTAGTCGAAGTACAATCCAGTATATCACACTTTAGAAAATTTCTCTGCT---GAAGAA 177
DB 117 ATCAGTGAAGATATATGCAAGCGCTTTCGCAATTTTGGAGTTTCTTTCGCAAGGAAGA 176
QY 178 ACTCCTTATTAATGAAACAATTTCTCTCACTAAAAAAGTTTTCGGACTAAAGAAAGATGGT 237
DB 177 AGAATAACAAGTTGGAGTATTTGGATAGACGACAGATTGGGATAGATGTAATATCT 236
QY 238 GATATAACAAGAGGAGTATTTTACAAGAGTAGCTCCAGGCAATGATTTTCAAAAATAAC 297
DB 237 AGAACCACTTTAAGCGATATATTCACCGTTTCCAAATATTTCATTTAAGTATGAAAAATA 296
QY 298 TTAATATCAGATTTTTCAGGAGTATTTGTTTACTCTATGAGCGGACCAAGAAATAGAACTT 357
DB 297 CTATTTTCAGGATTTGCAAGGAGCTATTGGCTTACTCAATTTGGATGGCCCAAGAAATAGAGCT 356
QY 358 GAAGCTGCATATCAACAATTTTAAATCCAAAAACACCGATAACAATGATGATCTGATAATG--- 415
DB 357 GAAGTATCTTATGAAGCATTCGATGTTTAAATCAAGGTAAACAATTAAGAACAAGCA 416
QY 416 -GTGAATACATAAACAATTTTGCATTTATCTCGTAAAGATGCAATGGAAGATCAGCAA--- 471
DB 417 CATAGATATTATGCTCTGTCTCCCTTCTCTCGGCACAGACACAGATAGATGGTGGCAGGC 476
```



472	Qy	-----TATGTAGTACTTAAAAATGACGGCACTACTTTTATGTCAATTGATGGTTAACT	525
477	Db	AGTGGCTCTGTCTTCTTAATAAATGAAGGACTACTTGATAAATCATTTATGCTCAACGCA	536
526	Qy	TGCTATGACATTACAGCTGAAGGAGTACTTTTCGTACCATATGCATGCGAGGTATAGGA	585
537	Db	TGTTATGATGTAATAAGTGAAGGCATACCTTTTCTCTTAATATGTGCAAGTATTGGT	596
586	Qy	GCAGATCTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGAAAAATA	645
597	Db	ATTGATTTAGTATCCATGTTTGAAGCTATAAATCCTAAATTTCTTATCAAGGAAATTA	656
646	Qy	GGTATTAGTTACCGTATCACACGAAGTCTCTGCAATTTATTTGGTGGATACTACCAATGC	705
657	Db	GGCTTAAGTTACCGTATTAAGCCACAGAAGCTCTCTGTTTTATTTGGTGGACATTTTTCATAAG	716
706	Qy	GTATTATGGTAATAAATTTGAGAAAGATACCTGTAATACTCCTGTAGTATTTAAATGATGCT	765
717	Db	GTGATAGGAACGAATTTAGAGATATTCCTACTATGATACCTAGTGAATCAGCGCTTGA	776
766	Qy	CCTCA---AACCACACTCTGCTTCAGTAACTCTTGACGTTGGATACTTTTGGCGGAGAAATT	822
777	Db	GGAAAAGGAACTACCCTGCAATAGTATACACTGACGTTGTTACTTTTGGCATAGAACTT	836
823	Qy	GGAAATGAGGTTCCACCTTC	840
837	Db	GGAGGAAGCTTTAACTTC	854

```

RESULT 26
US-10-059-964-7
; Sequence 7, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Orasi, No. US20020120115A1io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; CURRENT FILING DATE: 2002-01-28
; EARLIER APPLICATION NUMBER: 09/314,701
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(861)
US-10-059-964-7

```

Qy	238	GATATACAAAAAGACGATTTTCAAGAGTAGCTCAGGCATTGATTTTCAAAATAAC	239
Db	235	AGAACCACTTTAAGCGATATATTACCGTCCAAATTATTCATTTTAAGTATGAAAATAAT	294
Qy	298	TTAATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGGACGGACCAAGAATAGAACTT	357
Db	295	CTATTTTCAGGATTTGCAGGAGCTATTGGCTACTCTCAATGGATGGCCCAAGNATAGNCTT	354
Qy	358	GAAGCTGCATATCAACAATTTTAATCCAAAAAACCCGATACAAATGATCTGATTAATG--	415
Db	355	GAAGTATCTTATGAAGCATTCGATGTTAAAAATCAAGGTAACAAATTATAAGAACGAAGCA	414
Qy	416	-GTGAATACTATAAACATTTTGGCATATTCTCGTAAAGATGCAATGGAAGATCAGCAA---	471
Db	415	CATAGATATTATTGCTCTGTCGCCATCTTCTCGCACAGAGACACAGATAGATGGTGCAGGC	474
Qy	472	-----TATGTAGTACTTTAAAAATGACGGCATAACTTTTATGTCTCATGTGTTAATACT	525
Db	475	AGTGGCTGTCTTCTTAATAAATGAAGGACTACTTGTATAATCATTTATGCTGAACGCA	534
Qy	526	TGCTATGACATTACAGCTGAAGGAGTATCTTTTGGTACCATATGCAATGTGCGAGGTATAGGA	585
Db	535	TGTTATGATGTAATAAGTAGGAAGGCATACCTTTTCTCCTTATATATGTGCGAGGTATTGTT	594
Qy	586	GCAGATCTTATCAGTATTTTAAAGACCTCAATCTPAAATTTGCTTACCACGAAGAAAATA	645
Db	595	ATTGATTTAGTATCCATGTTTGAAGCTATAAATCCTAANAATTTCTTATCAAGGAANAATTA	654
Qy	646	GGTATTAGTTACCCCTATCACACCAGAAGTCTCTGCATTTATTGGTGGATPACTACCATGGC	705
Db	655	GGCTTAAGTTACCCTATAAAGCCAGAGCTTCTGTGTTTATTGGTGGACATTTTTCATTAAG	714
Qy	706	GTTATTGGTAAATAATTTGAGAAGNATACCTGTAATAACTCTCTGTAGTATTAATGATGCT	765
Db	715	GTGATAGGAACGAATTTAGAGATATTCCCTACTATGATACCCTAGTGAATTCAGCGCTTGA	774
Qy	766	CCTCA---AACCAATCTGCTTTCAGTAACTCTTGACGTTGGATACTTTTGGCGGAGAAATT	822
Db	775	GGAAGGAAGAACTACCCTGCAATAGTAACACTGGNCGTGTCTACTTTTGGCATAGNACTT	834
Qy	823	GGAATAGGTTTCACTTTC	840
Db	835	GGAGGAAGGTTTAACTTTC	852

RESULT 27  
US-10-062-994-8  
; Sequence 8, Application US/10062994  
; Patent No. US20020132789A1  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Bowie, Michael V.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Mahan, Suman M.  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Moreland, Annie L.  
; APPLICANT: Simbi, Bigboy H.  
; APPLICANT: Whitmire, William M.  
; APPLICANT: Alleman, Arthur R.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and  
; FILE REFERENCE: UP-167XC3  
; CURRENT APPLICATION NUMBER: US/10/062,994  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: 09/533,662  
; PRIOR FILING DATE: 2000-04-21  
; PRIOR APPLICATION NUMBER: 09/337,827  
; PRIOR FILING DATE: 1999-06-22  
; PRIOR APPLICATION NUMBER: 08/953,326  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 08/733,230

```
; PRIOR FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
US-10-062-994-8

Query Match      25.3%; Score 212.8; DB 12; Length 861;
Best Local Similarity 57.3%; Pred. No. 7.8e-36;
Matches 492; Conservative 0; Mismatches 342; Indels 24; Gaps 5;

QY 1 ATGAATTATAGAAAAATCTAGTAAGACGCGGTTAACTCTATTAATGTCAATCTTACCA 60
DB 3 ATGAACCTGCGAAAAATTTTATACAACTGATTAACATTAATTAATGTCTTCTTACCT 62
QY 61 TATCAGTCTTTTGCAGATCTCTAGGTTCAAGAACTTAATGATAACAAGAGGCTTCTAC 120
DB 63 GGAATATACATTTCTGTATCCAGTACAG-----GATGACAACTTAGTGGTAATTTCTAC 116
QY 121 ATTAGTGCAGAACTACAATCAAGTATATCACACTTTAGAAAAATTTCTGTCT---GAAGAA 177
DB 117 ATCAGTGGAAAGTATATGCAAGCGCTTCGCATTTTGGAGTTTTCCTGCCAAGGAAGAA 176
QY 178 ACTCCTATTATGGAACAAATTTCTCTACTAAAAAGTTTTCGGACTAAAGAAAGATGGT 237
DB 177 AGAAATACAAACAGTTGGAGTATTTGGATAGAGCAAGATTGGGATAGATGTGTAATATCT 236
QY 238 GATATAACAAAAAGACGATTTTACAGAGTAGCTCCAGGCAATTTGATTTTCAAAATAAC 297
DB 237 AGAACCACTTTAAGCGATATATTCACCGTTCCAAATTTATTTAAGTATGAAATATAT 296
QY 298 TTAATATCAGGATTTTCAGAAAGTATTTGGTTACTCTATGCGGACCAAGAAATAGAACTT 357
DB 297 CTATTTTCAGGATTTGCGAGGACTATTGGCTACTCAATGGATGGCCCAAGAAATAGAGCTT 356
QY 358 GAAGCTGCATATCAACAAATTTTAAATCCAAAAACACCGATAACAATGATCTATATG--- 415
DB 357 GAAGTATCTTTATGAAGCATTCGATGTTAAAAATCAAGGTAAACAAATTAAGAAGCAAGCA 416
QY 416 -GTGAATACTATAAACAATTTTGCATTTATCTCGTAAAGATGCAATGGAAGTACAGCAA--- 471
DB 417 CATAGATATTATGCTCTGTCGCCATCTTCTCGGCAACAGACACAGATAGATGGTCAGGC 476
QY 472 -----TAGTAGTACTTAAAAATGACGCGATAAATTTTATGTCATTTGATGGTTAATACT 525
DB 477 AGTGGGCTGCTCTTTCTAATAATGAAGGACTACTTGTATAAATCATTTATGCTGAACGCA 536
QY 526 TGCTATGACATTTACAGCTGAAGGATATCTTTTCGTACCATATGCAATGCGAGGTATAGGA 585
DB 537 TGTATGATGTAATAAGTGAAGGCAATCTTTCTCTTATATATGTCAGGATTTGGT 596
QY 586 GCAGATCTTATCACTATTTTAAAGACCTCAATCTTAAATTTTGCCTTACCAGGAAAAATA 645
DB 597 ATTGATTAGTATCATGTTGAGCTATAAATCTTAAATTTCTTATCAGGAAAAATTA 656
QY 646 GGTATAGTACCTATACACAGAAAGTCTCTGCAATTTATTTGGTGGATPACTACCATGGC 705
DB 657 GGCCTAAGTTACCTATAAGCCAGAGCTTCTGTGTTTATTTGGTGGACATTTTCATAAG 716
QY 706 GTTATTGGTAATAATTTGAGAGATACCTGTAATTAATCTCTGTAGTATTAATGATGCT 765
DB 717 GTGATAGGAAACGAATTTAGAGATATTTCTTACTATGATACCTAGTGAATCAGCGTTGCA 776
QY 766 CCTCA---AACACACTCTCTTTCAGTAACTCTTTGACGTTGGGATACCTTTGGCGGAGAAAT 822
DB 777 GGAAGGAAACCTACCCCTGCAATAGTAACACTGGAGCTGTTCTACTTTGGCATAGACTT 836
QY 823 GGAATGAGGTTCCACCTTC 840
DB 837 GGAGGAAGGTTTAACTTC 854
```

652 AGTTACCTTATCACACGAGAGTCTCTGCATTTATTTGGTGGATCTACTACCATGGCGTTATT 711  
655 AGTTACTTCCATAGCCAGAGAGTCTCTGTTTGGTGGACATTTTTCATAGGTGATA 714  
712 GGTATTAATTTGAGAGATACCTGTATTAATCTCTGTAGTATTAAATGATGCTCTCTCAA 771  
715 GGGATGAATTCAGAGATATTTCTTGGTGTATGATACCCAGTAGCTCAACTCTCACAGGTAAT 774  
772 ACCACATCTGCTTCAGTAACCTTTCAGCTTGGATCTTTGGCGGAGAAATTTGGAAATGAGG 831  
775 CAC---TTTACTATAGTAACACTAAGTGTATGCCACTTTGGAGTGGAACTTTGGAGGAGG 831  
832 TTCACCTT 839  
832 TTAACTT 839

RESULT 29  
US-10-062-994-10  
; Sequence 10, Application US/10062994  
; Patent No. US20010001661A1  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Bowie, Michael V.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Mahan, Suman M.  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Moreland, Annie L.  
; APPLICANT: Simbi, Bigboy H.  
; APPLICANT: Whitmore, William M.  
; APPLICANT: Allemen, Arthur R.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and  
; FILE REFERENCE: UF-167XC3  
; CURRENT APPLICATION NUMBER: US/10/062,994  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: 09/533,662  
; PRIOR FILING DATE: 2000-04-21  
; PRIOR APPLICATION NUMBER: 09/337,827  
; PRIOR FILING DATE: 1999-06-22  
; PRIOR APPLICATION NUMBER: 08/953,326  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 08/733,230  
; PRIOR FILING DATE: 1996-10-17  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 843  
; TYPE: DNA  
; ORGANISM: Ehrlichia chaffeensis  
US-10-062-994-10

Query Match 24.7%; Score 207.2; DB 12; Length 843;  
Best Local Similarity 56.2%; Pred. No. 1.2e-34;  
Matches 477; Conservative 0; Mismatches 353; Indels 18; Gaps 4;

1 ATGAATTAAGAAAATCTAGTAAGCGCGTTAATCTCAATTAATGCTCAATCTTACCA 60  
3 ATGAATTAAGAAAATTTTATTAACAACTACATAGTAGTATGCTCAATGCTCTTACCT 62  
61 TATCAGTCTTTTGGAGATCTGTAGGTTCAAGAACTAATGATAAAGAGGCTTCTAC 120  
63 GGAATATCAATTTCTGATGAGTACAGNACGACATGTTGGTGTAT-----TTCAT 116  
121 ATTAGTCAAGATCAATCCAAAGTATATCACTTTTGAAGAAATCTCTGTGAA---GAA 177  
117 ATCAGTGGGAATATGATACCAAGTGTTCACATTTTGGCGTATTCTCTGCTAAACAGAA 176  
178 ACTCTATTATGAACAAATCTCTCACTAAAAAGTTTTCGCACTAAGAAAGATGGT 237  
177 AGAAATCAACATCGGAGTATTTGGATTAAAGCAAGATTGGGATGGCAGCACATATCT 236

Qy 238 GATATAACAAAAAAGACGATTTTACAAGAGTAGTCCAGGCAFTGATTTTCAAAAAAATAC 297  
Db 237 AAAAAATCTCCAGAAAAATACATTTTAACGTTTCCAAATATTATTCATTTAAATATGAATAAT 296  
Qy 298 TTAATATCAGATTTTTCAGGAAGTATTTGGTGTACTCTATGACGACCAAGAAATAGAATT 357  
Db 297 CCATTTCTAGGTTTTCAGGAGCTGTTGGTTATTATTAATGAATGGTCCAAAGAAATAGAGTTA 356  
Qy 358 GAAGCTCATATCAACAATTTTAATCCAAAAACACCGATTAACATGATGATCTGATAATGGT 417  
Db 357 GAATGTCTTATGAAACATTTGATGGAAGAACACAGGATTAATTAATTAAGAACATGCT 416  
Qy 418 GAATATCTATAAACAATTTTGCATTTATCTCGT-----AAAGATGCAATGGAAGATCAGCAA 471  
Db 417 CACAAATATTATGCTTTTAAACCCATAACAGTGGGGGAAAGCTAAGCAATGCAGGTGATAAG 476  
Qy 472 TATGTAGTACTTAAAAATGACGCAATAACTTTTATGTCATTTGATGGTAAATGATCTGCTAT 531  
Db 477 TTTGTTTTCTTAAAAAATGAAGGACTACTTTGATATATACCTTATGTTTGAATGCAATGCTAT 536  
Qy 532 GACATTACAGCTGAAGAGTATCTTTCTGATACCATATGATGTCAGGTATAGGAGCAGAT 591  
Db 537 GATGTAATAAGTGAAGGAATACCTTTCTCTCTTACATATGTCAGGTGTTGGTACTGAT 596  
Qy 592 CTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAGGAAAAATAGGTATT 651  
Db 597 TTAATATCCATGTTTGAAGCTATAAACCCCTAAAAATTTCTTATCAAGGAAAGTTAGGTTT 656  
Qy 652 AGTTACCTATCACACAGAGTCTCTGCAATTTTATGTTGGTGGATACCTACCATGGCGTTATT 711  
Db 657 AGTTACTCCATAAGCCCAAGAGCTCTGTTTGTGTTGGTGGACATTTTTCATAAGGTGATA 716  
Qy 712 GGTAAATAAATTTGAGAGATACCTGTAATAACTCTGTAGTATTAAATGATGCTCCTCAA 771  
Db 717 GGGATGAATTCAGAGATATTTCTGCTATGATACCCAGTACCTCACTCTCACAGGTAT 776  
Qy 772 ACCACATCTGCTTTCAGTAACCTTTGAGGTTGGATCTTTGCGGAGAAATTTGGAATGAGG 831  
Db 777 CAC---TTTACTATGATTAACACTAAGTGTATGCCACTTTTGGAGTGGAACTTTGGAGGAAG 833  
Qy 832 TTCACCTT 839  
Db 834 TTAACTT 841

RESULT 30  
US-10-059-964-11  
; Sequence 11, Application US/10059964  
; Patent No. US20020120115A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20020120115A110  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; TITLE OF INVENTION: Chaffeensis  
; FILE REFERENCES: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/059,964  
; CURRENT FILING DATE: 2002-01-28  
; EARLIER APPLICATION NUMBER: 09/314,701  
; EARLIER FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 843  
; TYPE: DNA  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(843)  
US-10-059-964-11

Query Match 24.7%; Score 207.2; DB 12; Length 843;  
Best Local Similarity 56.2%; Pred. No. 1.2e-34;







Db 775 ---GCAATAGTAACTAAGTGTATGTATCTTTTGGAAATAGAACTTGGAGGAAGGTTTAAAC 831

Qy 838 TTC 840  
|||  
Db 832 TTC 834

## RESULT 35

US-10-062-994-9

; Sequence 9, Application US/10062994

; Patent No. US20020132789A1

; GENERAL INFORMATION:

; APPLICANT: Barbet, Anthony F.

; APPLICANT: Bowie, Michael V.

; APPLICANT: Burridge, Michael J.

; APPLICANT: Mahan, Suman M.

; APPLICANT: McGuire, Travis C.

; APPLICANT: Rurangirwa, Fred R.

; APPLICANT: Moreland, Annie L.

; APPLICANT: Simbi, Bigboy H.

; APPLICANT: Whitmire, William M.

; APPLICANT: Alleman, Arthur R.

; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and

; FILE REFERENCE: UP-167XC3

; CURRENT APPLICATION NUMBER: US/10/062,994

; PRIOR FILING DATE: 2002-01-31

; PRIOR APPLICATION NUMBER: 09/533,662

; PRIOR FILING DATE: 2000-04-21

; PRIOR APPLICATION NUMBER: 09/337,827

; PRIOR FILING DATE: 1999-06-22

; PRIOR APPLICATION NUMBER: 08/953,326

; PRIOR FILING DATE: 1997-10-17

; PRIOR APPLICATION NUMBER: 08/733,230

; PRIOR FILING DATE: 1996-10-17

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 9

; LENGTH: 837

; TYPE: DNA

; ORGANISM: Ehrlichia chaffeensis

US-10-062-994-9

Query Match 24.4%; Score 205; DB 12; Length 837;

Best Local Similarity 55.3%; Pred. No. 3.5e-34;

Matches 466; Conservative 0; Mismatches 365; Indels 12; Gaps 3;

Qy 1 ATGAATTATAGAAAATCTAGTAGAGCGGTTAATCTCAATTAATGTCATCTTACCA 60

Db 3 ATGAATTGCAAAAATTTTATACCACTGCATGTATCATCAATGTCCTTTCTACCT 62

Qy 61 TATCAGTCTTTTGCAGATCTGTAGTTTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120

Db 63 GGAATACATTTTCTGATCCAGTCA-----AGTGCATATATAGTGGTAATTTCTAT 116

Qy 121 ATTAGTCAAAAGTCAATCCAAAGTATATACATCTTTAGAAAATTTCTGCTGGAAGAACT 180

Db 117 GTTAGTGCAGATATATGCCAAGTGTGCGATTTTGGCATGTTTCTGCCAAAGAGAA 176

Qy 181 CCTATTATGAACAATTTCTCACTAATAAAGTTTTCGACTTAAGAAAGATGGTGAT 240

Db 177 AAAAACTCTACTGTGTCATGTTATGGCTTAAACAAGATTTGGGAAGGATTAGCTCATCA 236

Qy 241 ATAACAAAAAAGACGATTTTACAGAGTAGCTCCAGGCATTTGATTTTCAAAATAACTTA 300

Db 237 AGTCACATGATATATCAATTCATTAACACAGGGTTATTCATTTAATATGAATAACCCA 296

Qy 301 ATATCAGATTTTTCAGGAAGTATTGGTTACTTCTATGGACGCAACCAAGATAGAACTTGAA 360

Db 297 TTTTATAGGTTTTCAGGAGCTATTGGTTATTCAATGGTGTGCTCCCAAGAGTAGAGTTGAA 356

Qy 361 GCTGCATATCAACAATTTATCCAAAAAACACCATTAACATGATCTATGATATGGTGAA 420

Db 357 GTGTCCTATGAAACATTTTGACGTTTAAAAATCAGGGTAAATTAACATAAAAAATGATGCTCAC 416

Qy 421 TACTATAAACACATTTTGCATTTATCTCGTAAAGATGCAATG---GAAGATCAGCAATATGTA 477

Db 417 AGATACTGTGCTTTAGGTCAACAGACACACAGCGGAATACCTAAACTAGTAATACGTA 476

Qy 478 GTACTTAAAAATGACGGCATAACTTTTATGTCATGATGGTTAATATCTTGTCTATGACATT 537

Db 477 CTGTTAAAAAGCGAAGGATTTGCTTGACATATCATTTATGCTTAAATGCTATGATGATA 536

Qy 538 ACAGCTGAAGGAGTATCTTTGCTACCATATGTCATGTCGAGGTATAGGACGAGATCTTATC 597

Db 537 ATAAACGAGCATACCTTTGCTCTCTTACATATGTGAGGTGTGGTACTGATTTAATA 596

Qy 598 ACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAAATAGGTATTAGTTAC 657

Db 597 TCCATGTTGAAGCTACAAATCTTAAATTTCTTACCAAGGAAAGTTAGGTCTAAGTTAC 656

Qy 658 CTTATCACACGAGAGTCTCTGCAATTTATTGGTGGATCTACCATGGGGTTATTGGTAAAT 717

Db 657 TCTATAAACCCAGAGGCTTCTGTATTTATTGGTGGACATTTTTCATAAGGTGATAGGAAAC 716

Qy 718 AATTTTCGAGAGATACCTGTAATAACTCTCTGTAGTATTAAATGATGCTCTCAAAACCACA 777

Db 717 GAATTTAGGACATTTCTCTCTGAAAGCATTTGTTAGGTGATGATGATGATGATGATGATGAT 776

Qy 778 TCTGCTTTCAGTAATCTTTGACGTTGGATCTTTGGCGGAGAAATTTGGAATGAGGTTCCACC 837

Db 777 ---GCAATAGTAACTAAGTGTATGTCATTTTGGAAATAGAACTTGGAGGAAGGTTTAAAC 833

Qy 838 TTC 840

Db 834 TTC 836

## RESULT 36

US-10-062-994-3

; Sequence 3, Application US/10062994

; Patent No. US20010001661A1

; GENERAL INFORMATION:

; APPLICANT: Barbet, Anthony F.

; APPLICANT: Bowie, Michael V.

; APPLICANT: Burridge, Michael J.

; APPLICANT: Mahan, Suman M.

; APPLICANT: McGuire, Travis C.

; APPLICANT: Rurangirwa, Fred R.

; APPLICANT: Moreland, Annie L.

; APPLICANT: Simbi, Bigboy H.

; APPLICANT: Whitmire, William M.

; APPLICANT: Alleman, Arthur R.

; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and

; FILE REFERENCE: UP-167XC3

; CURRENT APPLICATION NUMBER: US/10/062,994

; PRIOR FILING DATE: 2002-01-31

; PRIOR APPLICATION NUMBER: 09/533,662

; PRIOR FILING DATE: 2000-04-21

; PRIOR APPLICATION NUMBER: 09/337,827

; PRIOR FILING DATE: 1999-06-22

; PRIOR APPLICATION NUMBER: 08/953,326

; PRIOR FILING DATE: 1997-10-17

; PRIOR APPLICATION NUMBER: 08/733,230

; PRIOR FILING DATE: 1996-10-17

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 842

; TYPE: DNA

; ORGANISM: Ehrlichia chaffeensis

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(840)

US-10-062-994-3

Query Match 24.3%; Score 203.8; DB 12; Length 842;  
Best Local Similarity 57.8%; Pred. No. 6.3e-34;  
Matches 424; Conservative 0; Mismatches 297; Indels 12; Gaps 3;

115 TTCTACATTAGTGCAGAACTACCAATCCAAAGTATATACACACTTTAGAAAATTTCTGCT--- 171  
109 TTCTACATGAGTGGAAATACGATCCAGGCTTCGCAATTTGGAGTATTTCTGCTGAAG 168  
172 GAAGAACTCTTATTAATGGAAACAAATTTCTCTACTNAAAAAGTTTTCGGACTAAAGAAA 231  
169 GAAGAAAGAAATACAAAGTGGAGTGTGGAGTGAAGCAAAATTTGGACGGAAGCGCA 228  
232 GATGCTGATATACAAAAAGACAAATTTTACAAAGTAGCTCCAGGCAATGATTTTCAA 291  
229 ATATCCAACTCTCTCCCAACGATGATTTCTACTGTCTCAAAATTAATTCATTTAAATATGAA 288  
292 AATACTTTAATACAGGATTTTCAGGAAGTATTTGTTACTCTATGACGCGACCAAGATA 351  
289 AACACCCGTTTTAGGTTTTCAGAGGCTATTTGTTACTCAATGATGGTCCCAAGATA 348  
352 GAATTTGAAGCTGATATCAACAAATTTAATCCAAAAACACCGATACCAATGATGAT 411  
349 GAGCTTGAAGTATCTTATGAACATTTTGTATGTAATAAATCAAGGTAAACAATTAAGAAT 408  
412 AATGCTGATATACAAATTTTGCATTTCTCTGTAAGTAGCAATG-----GAAGAT 465  
409 GAAGCACATAGATATTTGCTCTATCCCAATCTCAGCAGCAGACATGAGTAGTCAAGT 468  
466 CAGCAATATGATGATTTTAAATAATCAGCGCATACCTTTATGTCATTTGATGTTAATACT 525  
469 AATAATTTTGTCTTCTAATAAATGAGGATTTCTTGACATATCAATTTATGCTGACGCA 528  
526 TGCTATGATATPACAGCTGAAGAGTATCTTTTCGATCAATATGATGCTGAGGTATAGGA 585  
529 TGCTATGATGATGAGGAGGATACCTTTTCTCTTATATATGCGAGGTATCGGT 588  
586 CGAGATCTTATCACTATTTTAAAGACCTCAATCTAATAATTTGCTTACAGGAAATA 645  
589 ACTGATTTAGTATCCATGTTTGAAGCTACAAATCTTAAATTTCTTCAAGGAAAGTTA 648  
646 GGTATTTAGTACCTTATCAGCAGAGTCTCTGCAATTTATGTTGGATGATCTTACCATGGC 705  
649 GGTATTTAGTACTTATAGCCAGAGGCTCTGTTTATGTTTGGTGGGCACTTTCTAAG 708  
706 GTTATGTAATAATTTGAGAGATAC---CTGTAATACTCTCTGATGATTAATAATGAT 762  
709 GTATAGGGAACGAATTTAGAGATATCTCTACTATAATACCTTACTGGATCAACACTTGA 768  
763 GCTCTCAACACATCTGCTTCACTACTCTTGGATGATGATGATGATGATGATGATGATGAT 822  
769 GAAAGGAAATACCTCTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 828  
823 GGAATGAGGTTCA 835  
829 GGAGGAAGTTTA 841

## RESULT 37

US-10-062-994-3  
; Sequence 3, Application US/10062994  
; Patent No. US20020132789A1  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Bowie, Michael V.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Mahan, Suman M.  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Moreland, Annie L.  
; APPLICANT: Simbi, Bigboy H.  
; APPLICANT: Whitmire, William M.  
; APPLICANT: Allemen, Arthur R.

;; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and  
;; FILE REFERENCE: Methods of use  
;; CURRENT APPLICATION NUMBER: US/10/062,994  
;; CURRENT FILING DATE: 2002-01-31  
;; PRIOR APPLICATION NUMBER: 09/533,662  
;; PRIOR FILING DATE: 2000-04-21  
;; PRIOR APPLICATION NUMBER: 09/337,827  
;; PRIOR FILING DATE: 1999-06-22  
;; PRIOR APPLICATION NUMBER: 08/953,326  
;; PRIOR FILING DATE: 1997-10-17  
;; PRIOR APPLICATION NUMBER: 08/733,230  
;; PRIOR FILING DATE: 1996-10-17  
;; NUMBER OF SEQ ID NOS: 34  
;; SOFTWARE: Patent in Ver. 2.0  
;; SEQ ID NO 3  
;; LENGTH: 842  
;; TYPE: DNA  
;; ORGANISM: Ehrlichia chaffeensis  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (1)..(840)  
US-10-062-994-3

Query Match 24.3%; Score 203.8; DB 12; Length 842;  
Best Local Similarity 57.8%; Pred. No. 6.3e-34;  
Matches 424; Conservative 0; Mismatches 297; Indels 12; Gaps 3;

115 TTCTACATTAGTGCAGAACTACCAATCCAAAGTATATACACACTTTAGAAAATTTCTGCT--- 171  
109 TTCTACATGAGTGGAAATACGATCCAGGCTTCGCAATTTGGAGTATTTCTGCTGAAG 168  
172 GAAGAACTCTTATTAATGGAAACAAATTTCTCTACTNAAAAAGTTTTCGGACTAAAGAAA 231  
169 GAAGAAAGAAATACAAAGTGGAGTGTGGAGTGAAGCAAAATTTGGACGGAAGCGCA 228  
232 GATGCTGATATACAAAAAGACGATTTTACAAAGTAGCTCCAGGCAATGATTTTCAA 291  
229 ATATCCAACTCTCTCCCAACGATGATTTCTACTGTCTCAAAATTAATTCATTTAAATATGAA 288  
292 AATACTTTAATPACAGGATTTTCAGGAAGTATTTGTTACTCTCTATGACGCGACCAAGATA 351  
289 AACACCCGTTTTAGGTTTTCAGAGGCTATTTGTTACTCAATGATGGTCCCAAGATA 348  
352 GAATTTGAAGCTGATATCAACAAATTTAATCCAAAAACACCGATACCAATGATGAT 411  
349 GAGCTTGAAGTATCTTATGAACATTTTGTATGTAATAAATCAAGGTAAACAATTTAAGAAT 408  
412 AATGCTGATATACAAATTTTGCATTTCTCTGTAAGTAGCAATG-----GAAGAT 465  
409 GAAGCACATAGATATTTGCTCTATCCCAATCTCAGCAGCAGACATGAGTAGTCAAGT 468  
466 CAGCAATATGATGATTTTAAATAATCAGCGCATACCTTTATGTCATTTGATGTTAATACT 525  
469 AATAATTTTGTCTTCTAATAAATGAGGATTTCTTGACATATCAATTTATGCTGACGCA 528  
526 TGCTATGATATPACAGCTGAAGAGTATCTTTTCGATCAATATGATGCTGAGGTATAGGA 585  
529 TGCTATGATGATGAGGAGGATACCTTTTCTCTTATATATGCGAGGTATCGGT 588  
586 CGAGATCTTATCACTATTTTAAAGACCTCAATCTAATAATTTGCTTACAGGAAATA 645  
589 ACTGATTTAGTATCCATGTTTGAAGCTACAAATCTTAAATTTCTTCAAGGAAAGTTA 648  
646 GGTATTTAGTACCTTATCAGCAGAGTCTCTGCAATTTATGTTGGATGATCTTACCATGGC 705  
649 GGTATTTAGTACTTATAGCCAGAGGCTCTGTTTATGTTTGGTGGGCACTTTCTAAG 708  
706 GTTATGTAATAATTTGAGAGATAC---CTGTAATACTCTCTGATGATTAATAATGAT 762  
709 GTATAGGGAACGAATTTAGAGATATCTCTACTATAATACCTTACTGGATCAACACTTGA 768  
763 GCTCTCAACACATCTGCTTCACTACTCTTGGATGATGATGATGATGATGATGATGATGAT 822



```
Db      769 GGAAGGAAATACCCCTGCAATAGTAATCTGGATGTATGCCACTTTGGGAATGAAATG 828
Qy      823 GGAATGAGGTTCA 835
Db      829 GGAGGAAGGTTTA 841

RESULT 38
US-10-314-639-5
; Sequence 5, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE OF INVENTION: Chaffeensis
; CURRENT APPLICATION NUMBER: US/10/314,639
; CURRENT FILING DATE: 2002-12-09
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 843
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
; NAME/KEY: CDS
; LOCATION: (1)..(843)
US-10-314-639-5

Query Match      23.6%; Score 198.4; DB 9; Length 843;
Best Local Similarity 54.7%; Pred. No. 8.7e-33;
Matches 463; Conservative 0; Mismatches 371; Indels 12; Gaps 3;

Qy      1 ATGAATTATAGAAAAATCTAGTAGAAGCGGTTAATCTCATTAATGTCAAATCTTACCA 60
Db      1 ATGAAGTCGAAAAAATTTTATATACCACTGCAATGGCAATGGCCAAATGCTTTCTTACCT 60

Qy      61 TATCAGTCTTTGCGAGTCTGTAGGTTCAAGAACTAATGATTAACAAGAGGCTTCTAC 120
Db      61 GGAATATTACTTTCTGAACCAAGT-----ACAAGATGACAGTGTGAGTGGCAATTTCTAT 114

Qy      121 ATTAGTCAAAAGTACAATCAAGTATATACACTTTTAGAAAAATCTCTGCTGGAAGAACT 180
Db      115 ATTAGTGGCAAGTACATGCCAAGTCTTCTCATTTTGGAGTTTCTCTGCCAAGAGAA 174

Qy      181 CCTATTATGGAACAAATCTCTCACTAAAAAGTTTTCGGAAGTAAAGAAAGATGGTGAT 240
Db      175 AAAAAATCCTACTGTCGGGTTGTATGGTTTGAACAAGATTGGAACGGTGTGTAGTGTCTCA 234

Qy      241 ATACAAAAAAGACGATTTTACAGAGTAGTCCAGGCATTGATTTTCAAAATACTTA 300
Db      235 AGTCATGCTGATCGGACTTTTAAACAAGAGTTATCTTTTAAATACGAAAAACAATCCA 294

Qy      301 ATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGGACGACCAAGAAATAGAACTTGAA 360
Db      295 TTTCTAGGTTTTCAGGAGCTATTGGTTATTCAATGGGTGTCCAGAAATAGAGTTTGA 354

Qy      361 GCTGCATATCAACAAATTTAATCCAAAAACCAAGCAATATGATCTGATTAATGGTGA- 419
Db      355 GTGTCCTATGAACAAATTTGAGCTGAAAAATCAAGGTGGTAAATACAAAAATGATGCTCAC 414

Qy      420 --ATACTATAACATTTTGCATTATCTGTTAAAGATGCAATGGAAGATCAGCAATATGTA 477
Db      415 AGATATCTGTCCTTAGATCGTAAAGCAGACGACATAATGCCACAGCTAGTCACTACGTG 474

Qy      478 GTACTTAAAAATGACGGCAATACTTTTATGTCATTGATGGTTTAAATCTTGTATGACATT 537
Db      475 CTACTAAAAATGAGGACTACTTGATATATCACTTATGTTGAATGATGCTATGACGTA 534
```

```
Qy      538 ACAGCTGAAGAGATATCTTTGTTACCATATATGCTAGTGTGAGAGCAGATCTTATC 597
Db      535 GTAAGTGAAGCAATACCTTTCTCTCTTACATATATGTCAGGTGTGGTACCGATTTAATA 594

Qy      598 ACTATTTTAAAGACCTCAATCTAAATTTGCTTTACCAAGGAAAAATAGGTATTAGTTAC 657
Db      595 TCCATGTTTGAAGCTATAAACCCCTAAATTTCTTATCAAGGAAAGTTAGGTTTGGAGTTAC 654

Qy      658 CCTATCACACCAAGTCTCTGCAATTTATTGGTGGATACCTACCATGGCGTTATTGGTAAT 717
Db      655 TCTATAAACCCAGAAAGCTTCTGCTCTTTGGTGGACATTTTATATAAGTTGCAGGTAAT 714

Qy      718 AAATTTGAGAAGAT---ACCTGPAATACTCTCTAGTAGTATTAAATGATGCTCCTCAAAAC 774
Db      715 GAATTCAGGGGACATTTCTACTCTTAAAGCGTTTGTACACCATCATCTGCGAGTACTCCA 774

Qy      775 ACATCTGCTTCAGTAACCTCTTGACGTTGGATACCTTGGCGGAGAAATTTGGAATGAGGTTTC 834
Db      775 GACTTAGCAACAGTACACTGAGTGTGTGTCACCTTTGGAGTAGAACTTTGGAGGAAGATTT 834

Qy      835 ACCTTC 840
Db      835 AACTTC 840

RESULT 39
US-10-059-964-5
; Sequence 5, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; CURRENT FILING DATE: 2002-01-28
; EARLIER APPLICATION NUMBER: 09/314,701
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 843
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
; NAME/KEY: CDS
; LOCATION: (1)..(843)
US-10-059-964-5

Query Match      23.6%; Score 198.4; DB 12; Length 843;
Best Local Similarity 54.7%; Pred. No. 8.7e-33;
Matches 463; Conservative 0; Mismatches 371; Indels 12; Gaps 3;

Qy      1 ATGAATTATAGAAAAATCTAGTAGAAGCGGTTAATCTCATTAATGTCAAATCTTACCA 60
Db      1 ATGAAGTCGAAAAAATTTTATATACCACTGCAATGGCAATGGCCAAATGCTTTCTTACCT 60

Qy      61 TATCAGTCTTTGCGAGTCTGTAGGTTCAAGAACTAATGATTAACAAGAGGCTTCTAC 120
Db      61 GGAATATTACTTTCTGAACCAAGT-----ACAAGATGACAGTGTGAGTGGCAATTTCTAT 114

Qy      121 ATTAGTCAAAAGTACAATCAAGTATATACACTTTTAGAAAAATCTCTGCTGGAAGAACT 180
Db      115 ATTAGTGGCAAGTACATGCCAAGTCTTCTCATTTTGGAGTTTCTCTGCCAAGAGAA 174

Qy      181 CCTATTATGGAACAAATCTCTCACTAAAAAGTTTTCGGAAGTAAAGAAAGATGGTGAT 240
Db      175 AAAAAATCCTACTGTCGGGTTGTATGGTTTGAACAAGATTGGAACGGTGTGTAGTGTCTCA 234

Qy      241 ATACAAAAAAGACGATTTTACAGAGTAGTCCAGGCATTGATTTTCAAAATACTTA 300
Db      235 AGTCATGCTGATCGGACTTTTAAACAAGAGTTATCTTTTAAATACGAAAAACAATCCA 294

Qy      295 TTTCTAGGTTTTCAGGAGCTATTGGTTATTCAATGGGTGTCCAGAAATAGAGTTTGA 354
Db      301 ATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGGACGACCAAGAAATAGAACTTGAA 360

Qy      361 GCTGCATATCAACAAATTTAATCCAAAAACCAAGCAATATGATCTGATTAATGGTGA- 419
Db      355 GTGTCCTATGAACAAATTTGAGCTGAAAAATCAAGGTGGTAAATACAAAAATGATGCTCAC 414

Qy      420 --ATACTATAACATTTTGCATTATCTGTTAAAGATGCAATGGAAGATCAGCAATATGTA 477
Db      415 AGATATCTGTCCTTAGATCGTAAAGCAGACGACATAATGCCACAGCTAGTCACTACGTG 474

Qy      478 GTACTTAAAAATGACGGCAATACTTTTATGTCATTGATGGTTTAAATCTTGTATGACATT 537
Db      475 CTACTAAAAATGAGGACTACTTGATATATCACTTATGTTGAATGATGCTATGACGTA 534
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 00:12:31 ; Search time 1088 Seconds  
(without alignments)  
12503.868 Million cell updates/sec

Title: US-10-062-624-41  
Perfect score: 840  
Sequence: 1 atgaattataaagaattctt.....ttggaatgaggttcaccttc 840

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_nam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68.6	8.2	1101	17	CNS00EVL
2	63.4	7.5	750	17	CNS011ID
3	59.2	7.0	895	17	AZ550258
4	58.2	6.9	875	17	AZ550139
5	56.8	6.8	927	17	AZ549396
6	56.8	6.8	937	17	BH155225

c 7	55.8	6.6	889	17	AZ547516
c 8	55.8	6.6	1101	17	CNS0039G
c 9	55	6.5	1101	17	CNS0039G
c 10	54.2	6.5	639	17	CNS0170D
c 11	54.2	6.5	939	17	BH133767
c 12	53.6	6.4	897	17	AZ551013
c 13	53	6.3	1101	17	CNS01706
c 14	52.8	6.3	895	17	BH163498
c 15	52.8	6.3	922	17	BH136172
c 16	52	6.2	841	17	AZ689408
c 17	51.8	6.2	884	17	CNS0129A
c 18	51.6	6.1	606	13	BJ441163
c 19	51.6	6.1	819	17	AZ677439
c 20	51.6	6.1	833	17	AZ544101
c 21	51.6	6.1	912	17	AZ545164
c 22	51.6	6.1	912	17	BH158221
c 23	51.6	6.1	917	17	BH166533
c 24	51.6	6.1	928	17	BH144168
c 25	51.4	6.1	878	17	CNS0187R
c 26	51.2	6.1	884	17	AZ669586
c 27	51	6.1	932	17	AZ538693
c 28	51	6.1	932	17	BH132524
c 29	50.8	6.0	1001	17	CNS0155H
c 30	50.6	6.0	747	17	CNS011RQ
c 31	50.4	6.0	649	13	BM160056
c 32	50.4	6.0	852	17	AZ529741
c 33	50	6.0	481	13	B1323673
c 34	50	6.0	885	17	BH167558
c 35	50	6.0	886	17	BH146655
c 36	50	6.0	908	17	BH152307
c 37	50	6.0	936	17	BH160886
c 38	50	6.0	1201	17	CNS0167M
c 39	49.8	5.9	560	13	B1815836
c 40	49.8	5.9	868	17	AZ543954
c 41	49.8	5.9	891	17	AZ530903
c 42	49.6	5.9	484	13	BM146655
c 43	49.6	5.9	485	13	BM160769
c 44	49.6	5.9	557	13	BM169815
c 45	49.4	5.9	581	17	CNS01UP2

#### ALIGNMENTS

RESULT 1  
CNS00EVL  
LOCUS  
DEFINITION  
Drosophila melanogaster genome survey sequence T7 end of BAC:  
BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999  
Drosophila melanogaster genome survey sequence T7 end of BAC:  
BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
AL069706  
GI:4949849  
GSS.  
Drosophila melanogaster.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqret@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oseegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial



```

ACCESSION      AZ550258
VERSION        AZ550258.1  GI:11175559
KEYWORDS       GSS.
SOURCE         Entamoeba histolytica.
ORGANISM       Entamoeba histolytica.
REFERENCE      Eukaryota; Entamoebidae; Entamoeba.
AUTHORS        Loftus,B., Van Aken,S. and Fraser,C.
TITLE          Determination of clone end sequences from Entamoeba histolytica
               HMI:IMSS sheared DNA library
JOURNAL        Unpublished (2000)
COMMENT        Contact: Brendan J Loftus
               Department of Eukaryotic Genomics
               The Institute for Genomic Research
               9712 Medical Center Dr., Rockville, MD 20850, USA
               Tel: 301 838 0208
               Fax: 301 838 3543
               Email: bjlloftus@igr.org
               Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
               DNA library
               Seq primer: M13-Forward
               Class: shotgun
               High quality sequence start: 15
               High quality sequence stop: 659.

FEATURES       source
               1..895
               /organism="Entamoeba histolytica"
               /strain="HMI:IMSS"
               /db_xref="taxon:5759"
               /clone_lib="Entamoeba histolytica Sheared DNA"
               /note="Vector: pHOS1; Site 1: Bst I; Constructed at The
               Institute for Genomic Research (TIGR), Rockville, MD.
               Genomic DNA isolated from broth cultures of E. histolytica
               using a method described by Clark and Diamond (Clark,
               C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
               method for isolate identification. Exp. Parasitol.
               77:450.). The DNA was mechanically sheared to give a
               tight size distribution (~2 kb). The v + i method used for
               the library construction is described in detail in Smith,
               H.O. and Venter, J.C. (Making small insert libraries for
               whole genome shotgun sequencing projects. In Genome
               Sequencing: A Practical Approach, eds. M. Vaudin and B.
               Barrell, Oxford University Press, 1999)."
```

BASE COUNT 232 a 126 c 72 g 465 t

```

Query Match      7.0%; Score 59.2; DB 17; Length 895;
Best Local Similarity 46.1%; Pred. No. 0.0016;
Matches 199; Conservative 0; Mismatches 233; Indels 0; Gaps 0;

Qy 60 ATATCAGTCTTTTGACAGATCCTGTAGGTTCAAGAACTAATGATAACAAGAGGCTTCTA 119
Db 823 ATATCAGCATTAAGAAATATCCCTTTAGGTTAGCCAAACAAACAAAACGGGGGAG 764
Qy 120 CATTAGTCCAAGTACATCCAGTATATCATCTTTAGAAATTTCTCTGTGGAAGAAC 179
Db 763 CAACCATGATGAAGAGAGAAATTTTACCACTTTATAAATGTTTGTCTTAAATATT 704
Qy 180 TCCTATTATGGACAAATTTCTCTACATAAAGATTTTCGGACTAAGAAGATGTGA 239
Db 703 TGAAGAAAATTAAGATGTCATCTTAAGAGAGAAATATAGAAAGTCTTAATACAGAA 644
Qy 240 TATAACAAAAAGACGATTTTACAAGAGTAGTCCAGGCATTTGATTTTCAAAATACTT 299
Db 643 TGGAAATTAAGTAATGATATATATGATATTAAGAAATATTAAGAAGACATTTAAAT 584
Qy 300 AATATCAGATTTTCAGGAAGTATTTGGTTATCTTATGGACGGACAAGAAATAGAACTTGA 359
Db 583 AAAGAATGAAATTAAGAGAGAAATCAATAGAAATTAATGGAGGAGAAATATAAACAATCA 524
Qy 360 AGCTGCATATCAACATTTTATCCAAAAACACCGATACATGATGATCTGATTAATGTGA 419
Db 523 ATTGAAGATAAAGAAATTAAGAAATTAATATCAAAATTAAGAAAGAGAAATAGA 464
```

BASE COUNT 439 a 81 c 139 g 216 t

```

Query Match      6.9%; Score 58.2; DB 17; Length 875;
Best Local Similarity 47.9%; Pred. No. 0.0026;
Matches 168; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

Qy 141 AAGTATATCATCTTTAGAAATTTCTCTGTGGAAGAACTCTTATTAATGAACAAATTC 200
Db 418 AAATTTAACACTTTATAAATGTTTGTCTTTGAAATTTATTGAAAGAAAATTAAGATGCA 477
Qy 201 TCTCACTAAAAAGTTTTTCGGACTAAGAAGATGGTGATATACAAACAAAGACGATTT 260
Db 478 ATTAAGAGAGAAATAATAGAGAAATTAATACAGAAATTTGAAATTAATAATGATAT 537
```

BASE COUNT 439 a 81 c 139 g 216 t

```

Qy 420 ATACTATAAACATTTTGCATTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTAGT 479
Db 463 AGAATTTGAAGTAAATGAAATATTTTAAAGAAGAGAAATGAAGAACAAAGAGATGTACA 404
Qy 480 ACTTAAAAATCA 491
Db 403 ATTTTAAATTA 392

RESULT 4
AZ550139
LOCUS      ENTFU47TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.
ACCESSION  AZ550139
VERSION    AZ550139.1  GI:11175440
KEYWORDS   GSS.
SOURCE     Entamoeba histolytica.
ORGANISM   Eukaryota; Entamoebidae; Entamoeba.
REFERENCE  1 (bases 1 to 875)
AUTHORS    Loftus,B., Van Aken,S. and Fraser,C.
TITLE      Determination of clone end sequences from Entamoeba histolytica
               HMI:IMSS sheared DNA library
JOURNAL    Unpublished (2000)
COMMENT    Contact: Brendan J Loftus
               Department of Eukaryotic Genomics
               The Institute for Genomic Research
               9712 Medical Center Dr., Rockville, MD 20850, USA
               Tel: 301 838 0208
               Fax: 301 838 3543
               Email: bjlloftus@igr.org
               Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
               DNA library
               Seq primer: M13-Reverse
               Class: shotgun
               High quality sequence start: 20
               High quality sequence stop: 793.

FEATURES       source
               1..875
               /organism="Entamoeba histolytica"
               /strain="HMI:IMSS"
               /db_xref="taxon:5759"
               /clone_lib="Entamoeba histolytica Sheared DNA"
               /note="Vector: pHOS1; Site 1: Bst I; Constructed at The
               Institute for Genomic Research (TIGR), Rockville, MD.
               Genomic DNA isolated from broth cultures of E. histolytica
               using a method described by Clark and Diamond (Clark,
               C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
               method for isolate identification. Exp. Parasitol.
               77:450.). The DNA was mechanically sheared to give a
               tight size distribution (~2 kb). The v + i method used for
               the library construction is described in detail in Smith,
               H.O. and Venter, J.C. (Making small insert libraries for
               whole genome shotgun sequencing projects. In Genome
               Sequencing: A Practical Approach, eds. M. Vaudin and B.
               Barrell, Oxford University Press, 1999)."
```



## ORIGIN

Query Match 6.8%; Score 56.8; DB 17; Length 937;  
 Best Local Similarity 47.7%; Pred. No. 0.0056;  
 Matches 166; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

Qy 141 AAGTATATCATCAGTTAGAAAATCTCTGCTGAGAAACCTCTATTATGAGAACAAATTC 200  
 |||||  
 Db 90 AATTAAACACTTTATTAATGTTTTGCTTTGAAATTTATTTGAAAGAAAATTTAAGAGGTCA 149  
 |||||  
 Qy 201 TCTCACTAAAAAGTTTTCCGACTAAAGAAAGATGGTGATATAACAAAAAAGACGATTT 260  
 |||||  
 Db 150 ACTAAAGAGAAATATAGAAAGATTATACATGAATTTGGATCAAGTAAATGATAT 209  
 |||||  
 Qy 261 TACAAGAGTAGCTCCAGCATTTGATTTTCAAAATAACTTTAATATCAGGATTTTCAGGAAG 320  
 |||||  
 Db 210 TAATGAATATAAAGAAATAATAAAGAGACATTTTAAATAAAGAATGAATTTAAAGAGA 269  
 |||||  
 Qy 321 TATTGGTTACTCTATGACGACCAAGNATAGAACTTGAAGCTGCATATCAACAATTAA 380  
 |||||  
 Db 270 AATTAAATAGAAATATGAGGAGGAAGAAATTAATAAATCAATGTCAAGATAAAGAAATAA 329  
 |||||  
 Qy 381 TCCAAAAAACCCGATACATGATCTACTGATAATGGTGAATACTATAAACATTTTGGCATT 440  
 |||||  
 Db 330 AGAATAATACTAAATGAACAAGAAAGAAAGAAATAGAGAGAGGAGGTAAATGAAAT 389  
 |||||  
 Qy 441 ATCTCGTAAAGATCAATGGAAGATCAGCAATATGTAGTACTTTAAAA 488  
 |||||  
 Db 390 AATTAAAGAGAAATGAAAGAAACAAAGAAAGATGTACAAATTTATTAA 437  
 |||||

## RESULT 7

AZ547516/c  
 LOCUS ENT547516.1 889 bp DNA linear GSS 14-NOV-2000  
 DEFINITION Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.

ACCESSION AZ547516.1 GI:11170298  
 VERSION GSS.  
 KEYWORDS Entamoeba histolytica.  
 SOURCE Entamoeba histolytica.  
 ORGANISM Eukaryota; Entamoebidae; Entamoeba.  
 REFERENCE 1 (bases 1 to 889)  
 AUTHORS Loftus, B., Van Aken, S. and Fraser, C.  
 TITLE Determination of clone end sequences from Entamoeba histolytica HMI:IMSS sheared DNA library  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: bjlloftus@tigr.org  
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library  
 Seq primer: M13-Forward  
 Class: shotgun  
 High quality sequence start: 17  
 High quality sequence stop: 797.  
 Location/Qualifiers  
 1. 889  
 /organism="Entamoeba histolytica"  
 /strain="HMI:IMSS"  
 /db\_xref="taxon:5759"  
 /clone\_lib="Entamoeba histolytica Sheared DNA"  
 /note="Vector: pHOSt1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a

FEATURES  
source

Genoscope - Centre National de Sequencage :  
 Submitted (02-JUN-1999) Genoscope (E-mail : [secref@genoscope.cns.fr](mailto:secref@genoscope.cns.fr))  
 Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Batell, Oxford University Press, 1999).

BASE COUNT 223 a 124 c 71 g 471 t  
 ORIGIN

Query Match 6.8%; Score 55.8; DB 17; Length 889;  
 Best Local Similarity 47.6%; Pred. No. 0.0095;  
 Matches 165; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

Qy 146 TATCACACTTTAGAAAATCTCTGCTGAGAAACCTCTATTATGAGAACAAATTCCTCA 205  
 |||||  
 Db 744 TTTAACACTTTATAAATGTTGTTGTTTGAATTTATTGAAGAAAATTAAGATGTCAACTTA 685  
 |||||  
 Qy 206 CTAAAAAGTTTTCCGACTAAAGAAAGATGGTGATATAACAAAAAAGACGATTTTACAA 265  
 |||||  
 Db 684 AAGAAGAAATTAATAGAAAGATTTAATACATGAATTTGGAATTTAAAGTAAATTAATATG 625  
 |||||  
 Qy 266 GAGTAGCTCCAGGCATTGATTTTCAAAATAACTTTAATATCAGGATTTTCAGGAAAGTATTG 325  
 |||||  
 Db 624 AATATAAGAAATTAATAAGAGACATTTAATAATAGAAATGAAATTAAGAGAAATCA 565  
 |||||  
 Qy 326 GTTACTCTATGAGCGGACCAAGAAATAGAACTTTGAAGCTGCATATCAACAATTTTAAATCCAA 385  
 |||||  
 Db 564 ATAGAATAATGGAGGAGAAATAATAAACAATCAATGGAAGATAATGATTTAAAGAGAA 505  
 |||||  
 Qy 386 AAAACACCGATACAAATGATCTGATTAATGGTGATCTATTAACATTTTGCATTTCTC 445  
 |||||  
 Db 504 TAATATCAAAAGTGAACAAGAAAGAAAGAAATAGAAAGATGAAGTAAATGAAATATTTA 445  
 |||||  
 Qy 446 GTAAAGATGCAATGGAAGATCAGCAATATGTAGTACTTTAAATAATGAC 492  
 |||||  
 Db 444 AAGAGAGAAATGAAGAGACAGAGATGTACATTTCAATTTACTAC 398  
 |||||

## RESULT 8

CNS0039G/c  
 LOCUS CNS0039G.1 1101 bp DNA linear GSS 03-JUN-1999  
 DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL063921  
 VERSION AL063921.1 GI:4941778  
 KEYWORDS GSS.  
 SOURCE Drosophila melanogaster.  
 ORGANISM Drosophila melanogaster.  
 REFERENCE 1 (bases 1 to 1101)  
 AUTHORS Drosophila melanogaster.  
 TITLE Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 JOURNAL Genoscope.

Direct Submission  
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : [secref@genoscope.cns.fr](mailto:secref@genoscope.cns.fr))  
 Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

1 (bases 1 to 1101)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazuhiro Ohsawa and  
Aaron Mamoser in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's  
P1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

Location/Qualifiers  
1..1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="BACR08K10"  
/clone\_lib="RPCI-98"  
/note="end : TET3"

BASE COUNT 201 a 64 c 131 g 202 t 503 others

ORIGIN

Query Match 6.5%; Score 55; DB 17; Length 1101;  
Best Local Similarity 19.6%; Pred. No. 0.015;  
Matches 118; Conservative 228; Mismatches 253; Indels 2; Gaps 1;

QY 171 TGAAGAACTCCTATTAATGAACAAATCTCTCACTAAAGAGTATGTTTCCGAGCTAAGAA 230  
DB 399 TAWAWWWWWTTTTTTTAAWAAAAAATAATTTWAAWAAAAAATAATTTWAAWAAAAA 458  
QY 231 AGATCGTATATACAAAAAAGACGATTTTACAAGAGTAGCTCCAGGCATTGATTTTCA 290  
DB 459 WTAWWTTTATWAAAAAATAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAA 518  
QY 291 AAATAACTTAATATCAGGATTTTTCAGGAAGTATGTTTCTCTATGACGACCAAGAAT 350  
DB 519 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 578  
QY 351 AGAATCTGAGCTGCATATCAACAAATTTAATCCAAAAACACCGATACCAATGATCTGA 410  
DB 579 TVHYTY 638  
QY 411 TAA--TGGTGAATCATATAACATTTTGCATATCTCGTAAAGATGCAATGGAAGATCAG 468  
DB 639 YYYTYTCYHYHYHHHHHA 698  
QY 469 CAATATGTAGTACTTAAATAATGACGACATACCTTTTATGTCAATGATGTTTAACTTGC 528  
DB 699 HHCYHY 758  
QY 529 TATGACATTAACAGCTGAAGAGTATCTTCTGTCATATGTCATGTCAGGTATAGGACGA 588  
DB 759 AWAHAAMWWHHAAWAAAAAATAATTTTHYHYHYHYHYHYHYHYHYHYHYHYHYHYHY 818  
QY 589 GATCTATACATTTTAAAGACCTCAATCTAAATTTTCTTACCAAGGAAATAAGT 648  
DB 819 YHTATCTWTTWTHWTHWTHWTHWTHWTHWTHWTHWTHWTHWTHWTHWTHWTHWTH 878  
QY 649 ATTAGTTACCTTATCACACCAAGAGTCTCTCATTTTATGTCATGATCACTACCTGCG 708  
DB 879 CMTMHH 938  
QY 709 ATTGGTAATAATTGAGAAGATCACTGCTTAATAACTCTCTAGTATTAATGATGCTCT 768

FEATURES  
source

Query Match 6.5%; Score 55.8; DB 17; Length 1101;  
Best Local Similarity 17.7%; Pred. No. 0.0097;  
Matches 112; Conservative 258; Mismatches 257; Indels 6; Gaps 1;

QY 1 ATGAATTAAGAAAATCTAGTAAAGCGGTTAAATCTCAATATGCTCAATCTTACCA 60  
DB 1080 WTKWTTWKDRAADRKAWGADRWADGAGTWTATWTTWWWWATWDTWDDKWWAT 1021  
QY 61 TATCAGTCTTTTCGAGATCTCTAGGTTCAGAACTAATGATTAACAAAGAGGCTTCTAC 120  
DB 1020 AAKTDTATWTTWRTAWRADWAGRDGAKGRDRDAATDAGAGRRDGRKDKDKDGD 961  
QY 121 ATTAGTGCAAGTACCAATCAAGTATATACACTTTAGAAATTTCTGCTGAAGAACT 180  
DB 960 DRKGGKKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 901  
QY 181 CCTATTAAATGGAACAATTTCTCACTAAAGATTTTCGGACTAAAGAAAGATGGTAT 240  
DB 900 DTGTTKDDDDKDKWDDWAKAGTGWGATWAAATDWWWWGADADWTTWDAADW 841  
QY 241 ATAACAAAAAGAGAGATTTTACAAGAGTAGCTCCAGGCATTGATTTTCAAAATAACTTA 300  
DB 840 RWDAAWAKWDDAWGARTADRDWDGAGKRGARKRRDRKRDADDDKRDADDDAATW 781  
QY 301 ATATC-----AGGATTTTCAGAGTATTTGGTTACTCTATGACGACGACCAAGATGAA 354  
DB 780 TTTTTRDTRDTRDTRDTRDTRDTRDTRDTRDTRDTRDTRDTRDTRDTRDTRDTR 721  
QY 355 CTGGAAGCTGCATATCAACAATTTTATCCAAAAACACGATTAACAATGATGATAAT 414  
DB 720 TRWDDADADDTARDDRRRRGGDAGDGKGTGKRRRRDRDRTDRTDRTDRTDRTDRT 661  
QY 415 GGTGAATACATATAAATTTTGCATTTATCGTAAAGATGCAATGGAAGATCAGCAATAT 474  
DB 660 DTDTDDDKDRRRKRRRRRTTAAAWDWWTWKAWDWWKAWDWWKAWDWWKAWDWW 601  
QY 475 GTAGTACTTAAATGACGACATCTTTTATGCTCAATGATGTTTAACTGCTATGAC 534  
DB 600 DARKADRWAKAWRARDRAARDRRTTGTGTTTATTTTAAWAAWAAWAAWAAW 541  
QY 535 ATTACAGCTGAAGAGTATCTTTCGATACATATGATGTCAGGTATAGGACGATCTT 594  
DB 540 TATWTTTWTWTTTWTWTTTWTWTTTWTWTTTWTWTTTWTWTTTWTWTTTWTWTT 481  
QY 595 ATCACTATTTTAAAGACCTCAATCTAAATTT 627  
DB 480 TTTTWTWTTTWTWTTTWTWTTTWTWTTTWTWTTTWTWTTTWTWTTTWTWTTTWT 448

CNS0039G 1101 bp DNA linear GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM



```

Db      939 HWWATWTTTMMCMCHHCHYHMHMTMYCCHYYCTCHTATTHYHMTCTYH 998
Qy      769 C 769
Db      999 Y 999

RESULT 10
CNS017QD      639 bp      DNA      linear      GSS 26-JUL-1999
LOCUS      Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION      BACN37M13 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION      ALI08367.1      GI:5628671
VERSION      GSS.
KEYWORDS      Drosophila melanogaster.
SOURCE      Drosophila melanogaster.
ORGANISM      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 639)
AUTHORS      Genoscope.
TITLE      Direct Submission
JOURNAL      Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT      Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

FEATURES             source
    source
    1..639
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /clone="BACN37M13"
        /clone_lib="DrosBAC"
        /plasmid="pBelobAC11"
        /note="end : SP6"
BASE COUNT      389 a      53 c      46 g      52 t      99 others
ORIGIN
Query Match      6.5%; Score 54.2; DB 17; Length 639;
Best Local Similarity 37.5%; Pred. No. 0.022;
Matches 119; Conservative 45; Mismatches 153; Indels 0; Gaps 0;

Qy      90 AAGAACTAATGATAACAAAGAGGCTTCTACATTAGTGCAGAAATGACAAATCCAAGTATATC 149
Db      178 AAGAAAAAATAAAAAAARAAGTAGGAGGTATTAAWTTAAAWAAAAAARAAWATA 237
Qy      150 ACACCTTAGAAAAATCTCTGCTGAAGAACTCTTATTAAATGGAACAAATCTCTCACTAA 209
Db      238 TCAACWAAAAAWARTAAWTAARAAARAAWATGAGTAARAAAWATWTAAWRTTTAAGAA 297
Qy      210 AAAAGTTTCCGACTAAGAAGATGTTGATATACAAARAAACAGCATTTTACAGAGT 269
Db      298 AAAAAAAGWGTGGAWWARARAAARAAAWAAWAAWAAATGAARARAAAGAWAGTGGAA 357
Qy      270 AGCTCCAGGCAATTCATTTTCAAAAATACTTAATATCAGGATTTTCAGGAAGTATTGGTTA 329
Db      358 AACACAAAWARAAARAAAGAAARAAAGAAWAAWAAWRAWRTWGAWGTAAWAAWRAAW 417
Qy      330 CTCTATGACGACCAAGAAATAGAACTTGAAGCTGCATATCAACAAATTTATCCAAAAA 389
Db      418 GTATTAAAAAATAAAAAAARAAARAAAWAAAAAACAACAAAGAAATGMAAAAAAATA 477
Qy      390 CACCGATAACCAATGATA 406
Db      478 AAAAAAATAAAAAAATA 494

```

```

RESULT 11
BH133767/c
LOCUS      ENT0237TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION      genomic, DNA sequence.
ACCESSION      BH133767.1      GI:15092828
VERSION      GSS.
KEYWORDS      Entamoeba histolytica.
SOURCE      Entamoeba histolytica.
ORGANISM      Eukaryota; Entamoebidae; Entamoeba.
REFERENCE      1 (bases 1 to 939)
AUTHORS      Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE      Determination of clone end sequences from Entamoeba histolytica
JOURNAL      HM1:IMSS sheared DNA library (2001)
COMMENT      Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 16
High quality sequence stop: 794.

```

```

FEATURES             source
    source
    1..939
        /organism="Entamoeba histolytica"
        /strain="HM1:IMSS"
        /db_xref="taxon:5759"
        /clone_lib="Entamoeba histolytica Sheared DNA"
        /note="Vector: PHOSI; Site 1: Bat 1; Constructed at The
        Institute for Genomic Research (TIGR), Rockville, MD.
        Genomic DNA isolated from broth cultures of E. histolytica
        using a method described by Clark and Diamond (Clark,
        C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
        method for isolate identification. Exp. Parasitol.
        77:450.). The DNA was mechanically sheared to give a
        tight size distribution (~2 kb). The v + i method used for
        the library construction is described in detail in Smith,
        H.O. and Venter, J.C. (Making small insert libraries for
        whole genome shotgun sequencing projects. In Genome
        Sequencing: A Practical Approach, eds. M. Vaudin and B.
        Barrel, Oxford University Press, 1999)."
```

```

BASE COUNT      224 a      135 c      75 g      505 t
ORIGIN
Query Match      6.5%; Score 54.2; DB 17; Length 939;
Best Local Similarity 47.3%; Pred. No. 0.022;
Matches 164; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

```

```

Qy      146 TATCACACTTTAGAAATCTCTGCTGAAGAACTCTATTATGGAACAAATCTCTCA 205
Db      806 TTTAAACATCTTATAAATGTTTGTCTTAAATATTGGAAGAAATTAAGAGGCTCACTAA 747
Qy      206 CTAAAAAAGTTTTCCGGACTAAAGAAAGATGGTGATATAACAAAAAGACGATTTTACAA 265
Db      746 AAGAGAAATAATAGAAGGTTAATACATGAATTTGAATTAATAATGATATTATG 687
Qy      266 GAGTAGTCCAGGCAATGATTTTCAAATTAATTAATATCAGGATTTTCAGGAAGTATTG 325
Db      686 AATATAAGAAATAATAAAGAAGATATTTAAATAAAGAATGAATTAATAAGATAAATCA 627
Qy      326 GTTACTCTATGGACCGACCAAGATAGAACTTGAGCTGCATATCAACATTTTATCCNA 385
Db      626 ATAGAATAATGGAGGAAGNAATAATAAACCACCAATGGAAGATAAAGAAATAAAGAA 567

```



890 GGATWAPAGACRRKAAATAGARAWAWAAGAAWAAAAAADDKABTDWATAAWAA 949  
 390 CACGATACCAATGATGATGTAATGTTGTAATCTATATAACATTTTGCAATCTCGTAA 449  
 950 AAAAAAGDWATKKTWATACATATRAVAATGKNGTTTTTTRTKAAAAAAGCCARADAANGTGT 1009  
 450 AGATGCAATGCAAGATCAGCATATGCTAGTACTTAAATAATCAGCGCATAACTTTTATGTC 509  
 1010 KGATGAAAAAAGWADGAVGATDADRAWRYKTATRTTABARTSTCWNYAKKDTNTATT 1069  
 510 ATTGATGCTTAATA 523  
 1070 AAKSARGCTAWWR 1083

RESULT 14  
 BH163498/c  
 LOCUS ENTREQ70TR Entamoeba histolytica Sheared DNA Entamoeba histolytica  
 DEFINITION genomic, DNA sequence.  
 ACCESSION BH163498  
 VERSION BH163498.1 GI:15736936  
 KEYWORDS GSS.  
 SOURCE Entamoeba histolytica.  
 ORGANISM Entamoeba histolytica.  
 Eukaryota; Entamoebidae; Entamoeba.  
 REFERENCE 1 (bases 1 to 895)  
 Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.  
 Determination of clone end sequences from Entamoeba histolytica  
 HMI:IMSS sheared DNA library (2001)  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: bjloftus@tigr.org  
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
 DNA library  
 Seq primer: M13-Reverse  
 Class: shotgun  
 High quality sequence start: 13  
 High quality sequence stop: 829.  
 Location/Qualifiers  
 FEATURES  
 source  
 1..895  
 /organism="Entamoeba histolytica"  
 /strain="HMI:IMSS"  
 /db\_xref="taxon:5759"  
 /clone\_lib="Entamoeba histolytica Sheared DNA"  
 /note="Vector: pHOS1; Site 1: Bst I; Constructed at The  
 Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from broth cultures of E. histolytica  
 using a method described by Clark and Diamond (Clark,  
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
 method for isolate identification. Exp. Parasitol.  
 77:450.). The DNA was mechanically sheared to give a  
 tight size distribution (~2 kb). The v + i method used for  
 the library construction is described in detail in Smith,  
 H.O. and Venter, J.C. (Making small insert libraries for  
 whole genome shotgun sequencing projects. In Genome  
 Sequencing: A Practical Approach, eds. M. Vaudin and B.  
 Barrell, Oxford University Press, 1999)."  
 223 a 117 c 83 g 472 t  
 BASE COUNT  
 ORIGIN  
 Query Match 6.3%; Score 52.8; DB 17; Length 895;  
 Best Local Similarity 46.9%; Pred. No. 0.047;  
 Matches 165; Conservative 0; Mismatches 187; Indels 0; Gaps 0;  
 141 AAGTATATCACATTTAGAAAATCTCTGCTGGAAGAACTCTTATTATGGAACAAATTC 200  
 654 AATCTAACACTCTATATAATGTTTGGCTTTAAATAATTATTGGAAGGAATTAAGATGTCA 595

201 TCTCACTAAAAAAGTTTTTCGACTTAAAGAAAGATGGTGATATAACAAAAAAGACGATTT 260  
 594 ACTTAAAGAGAGAAATAATAGAAAGATTAAATACAGGAATTTGGAAATTTAAAGTAAATGATAT 535  
 261 TACAAGAGTAGTCCAGGCATTGATTTTCAAAATAACTTAATATCAGGATTTTCAGGAG 320  
 534 TAATGAATATAAAGAAAGTAATAAAGAGACATTTTAAATAAAGAAATGAATGAATTTAAAGAGA 475  
 321 TATTGGTTTACTCTATGACGAGCACCAGAAATAGAACTTTGAAGCTCATATCAACAATTTAA 380  
 474 AATCAATAGATAATGAGTAAGAAATATAAACAATCAATGTCAGATATGATTAATTA 415  
 381 TCCAAAAAACACCGATTAACAATGATCTGTAATGTTGGAATATCTATATAACAATTTTGCATT 440  
 414 AGAAATAATATCAAAAGTGTAACAAGAAAGAAAGAAATAGNAGAGATGAAGTAATGAAAT 355  
 441 ATCTCGTAAGATCAATGGAAGATCAGCAATATGTAGTACTTAAATAATGAC 492  
 354 ATTTAAAGAAAGAAAGAAAGAAAGAAAGAAAGATGTCCCAATTTTATTAACACTAC 303

RESULT 15  
 BH136172  
 LOCUS ENTQX40TF Entamoeba histolytica Sheared DNA Entamoeba histolytica  
 DEFINITION genomic, DNA sequence.  
 ACCESSION BH136172  
 VERSION BH136172.1 GI:15095233  
 KEYWORDS GSS.  
 SOURCE Entamoeba histolytica.  
 ORGANISM Entamoeba histolytica.  
 Eukaryota; Entamoebidae; Entamoeba.  
 REFERENCE 1 (bases 1 to 922)  
 Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.  
 Determination of clone end sequences from Entamoeba histolytica  
 HMI:IMSS sheared DNA library (2001)  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: bjloftus@tigr.org  
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
 DNA library  
 Seq primer: M13-Forward  
 Class: shotgun  
 High quality sequence start: 6  
 High quality sequence stop: 484.  
 Location/Qualifiers  
 FEATURES  
 source  
 1..922  
 /organism="Entamoeba histolytica"  
 /strain="HMI:IMSS"  
 /db\_xref="taxon:5759"  
 /clone\_lib="Entamoeba histolytica Sheared DNA"  
 /note="Vector: pHOS1; Site 1: Bst I; Constructed at The  
 Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from broth cultures of E. histolytica  
 using a method described by Clark and Diamond (Clark,  
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
 method for isolate identification. Exp. Parasitol.  
 77:450.). The DNA was mechanically sheared to give a  
 tight size distribution (~2 kb). The v + i method used for  
 the library construction is described in detail in Smith,  
 H.O. and Venter, J.C. (Making small insert libraries for  
 whole genome shotgun sequencing projects. In Genome  
 Sequencing: A Practical Approach, eds. M. Vaudin and B.  
 Barrell, Oxford University Press, 1999)."  
 458 a 87 c 131 g 246 t  
 BASE COUNT  
 ORIGIN

	Query Match	6.2%;	Score 52;	DB 17;	Length 841;
	Best Local Similarity	46.8%;	Pred. No. 0.072;		
	Matches 163;	Conservative 0;	Mismatches 185;	Indels 0;	Gaps 0;
Qy	141	ACGTATATCACACTTTAGAAAAATTCCTGCTGAAGAAACTCCTATTAAATGGAACAAATTC	200		
Db	749	AAATTTACACTTTATAATGTTTCTCTTGGAAATTTATTGGAAGAAAATTAAGATGTCA	690		
Qy	201	TCTCACTAAAAAAGTTTTCCGGACTAAAGAAAGATGCTGATATAACAAAAAAGACGATTT	260		
Db	689	ATTAAAGGAAGAAATAATAGAAAGATTAAATACAAGAAATTTGAAATTTAAAGTAAATCATAT	630		
Qy	261	TACAAGATGAGTCCAGGCGATTTGATTTCCAAATAACTTAATATCAGGATTTTCAGGAAG	320		
Db	629	TATGTAATATAAGAAATAATAAAGAAAGACATTTAAATAAAGAAATGAAATAAAAAAGAGA	570		
Qy	321	TATTGGTTACTCTATCGGCGGCCAGAAATAGAACTTGAAGCTGCAATATCAACAAATTTAA	380		
Db	569	AATAAAATAGAATAATCGGGAAGAAATAATAAACATCAATGGAAAGATAAGTAAATTTAA	510		
Qy	381	TCCAAAAAACACCGGATAACAATGATCTGTAATGCTGAATACTATAAAACATTTTCGATT	440		
Db	509	AGAAATATAACCAAAATGNAACATAAAGAGAAATAGAAAGAGAGATATAATGAAAT	450		
Qy	441	ATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTAGTACTTTAAAAA	488		
Db	449	ATTTTAAAGGAAGAAATGAAGAAACAAGAAAGATGTACAATTTATTAA	402		

Db 509 AGAATAATAACCAAAATGAACAAATAAGAAGAAATAGAGAAGAGGATATATGAAAT 450

Qy 441 ATCTCGTAAAGTGCATGGAAGATCAGCAATATGTAGTACTTTAAAA 488

Db 449 ATTTAAAGAAAGAANAATGAAGACACAGAAGATGTCAATTTTATTA 402

Genoscope.  
Direct Submission  
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrs@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
library (Dros BAC) was made by Alain Billaut at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelOBAC11.

```

http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOAC11.

FEATURES             Location/Qualifiers
     source            1..884
                        /organism="Drosophila melanogaster"
                        /db_xref="taxon:7227"
                        /clone="BACN07H19"
                        /clone_lib="DrosBAC"
                        /plasmid="pBelOAC11"
                        /note="end : T7"
BASE COUNT            158 a   55 c   148 g   471 t   52 others

```

## ORIGIN

Query Match 6.2%; Score 51.8; DB 17; Length 884;  
Best Local Similarity 40.8%; Pred. No. 0.08;  
Matches 179; Conservative 24; Mismatches 236; Indels 0; Gaps 0;

QY 68 CTTTTCAGATCCTCTAGGTTCAAGAACTAATGATAACAAGAGAGCTTCTACATTAGTG 127  
DB 609 CTTGTCAGAGCCTCCGACAAAWAAATAWAAWAAAAAATAWAAWAAAAAATAWAAWAAAAAATA 550

QY 128 CAAAGTACATCCAAAGTATATCACACTTTAGAAAAATCTCTGCTCAAGAACTCTCTATTA 187  
DB 549 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 490

QY 188 ATGGAACTATCTCTCACTAAAAAGTTTTCGGCTAAAGAAAGATGGTGATATAACAA 247  
DB 489 TTAATAATAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 430

QY 248 AAAAAAGCATTTTACAGAGTAGCTCCAGCAATGATTTTCAAAATTAATTAATATCAG 307  
DB 429 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 370

QY 308 GATTTTCAGGAAGTATGCTTACTCTATGGAGCGACCAAGATAGAACTTGAAGCTGCAT 367  
DB 359 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 310

QY 368 ATCAACAATTTAATCCAAAAACCCGATAACAATGATAGTAAATGGTGAATACTATA 427  
DB 309 AACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 250

QY 428 AACATTTTGCATATCTCTGAAGATGCAATGGAAGATCAGCAATATGTAGTACTTAAAAA 487  
DB 249 AAAAAATWACAATTTTATATWAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 190

QY 488 ATGACGGCATAACTTTTAT 506  
DB 189 ATACATGTTACTTTTCAAT 171

RESULT 18  
BJ441163/c  
LOCUS  
DEFINITION BJ441163 Dictyostelium discoideum cDNA library, VF Dictyostelium  
discoideum cDNA clone ddv45124 3', mRNA sequence.

ACCESSION BJ441163.1 GI:19415885  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 606)  
Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium..  
Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.  
Full length cDNA of Dictyostelium discoideum at the vegetative stage  
Unpublished (2002)  
Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.

## FEATURES

source  
1..606  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone="ddv45124"  
/clone.lib="Dictyostelium discoideum cDNA library, VF"  
/sex="mat A"  
/dev\_stage="Growth phase"  
161 a -66 c 76 g 302 t 1 others

## BASE COUNT

## ORIGIN

Query Match 6.1%; Score 51.6; DB 13; Length 606;  
Best Local Similarity 46.2%; Pred. No. 0.087;  
Matches 171; Conservative 0; Mismatches 199; Indels 0; Gaps 0;

QY 247 AAAAAAGCAGATTTTACAGAGTAGCTCCAGGCATTTGATTTTCAAAAATAACTTAATATCA 306  
DB 571 AAAAAAACAACAATAATCCAAATCCAACTCAACAATTTGTTTCCCAAGTAAATTTATCAATT 512

QY 307 GGATTTTTCAGGAAGTATTTGTTTACTCTATGACGGACCAAGAATAAGAACTTTGAAGCTGCA 366  
DB 511 AATGTTTCAGTATCAAAATCATCAACATTAATGCTGTAATTAACCTGGAAGTTAA 452

QY 367 TATCAACAATTTAATCCAAAAAACAACCGATTAACATGATCTGATAATGGTGAATACTAT 426  
DB 451 CAACAAAAACAACAAGAAATTTAAAAATTGAACTGTAATTTAAAAAAGAAGAGAGATT 392

QY 427 AAACATTTTGCATTTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTAGTACTTAAA 486  
DB 391 AAAAAAGGTAGAACAGAGGTTTAAAAAAGTTTAAAGAACCAAGAGAACTTGAATTTATT 332

QY 487 AATGACGCGATAACTTTTATGTCATGATGTTGTTAATCTTCTGCTATGACATTACAGCTGAA 546  
DB 331 AATGAGGAACATGATGTTGTAATAATGGAGATTAATTTCAAACTTGAATAATCAA 272

QY 547 GGAGTATCTTTTCGTACCATATGTCATGTCAGGTATAGGAGCAGATCTTTATCACTATTTT 606  
DB 271 ATTAAGAACTTGAATACAGAAACAAATAAATAATATAATATTAATTAATTAATTAAT 212

QY 607 AAGACCTCA 616  
DB 211 AAAAAACAACA 202

## RESULT 19

AZ677439/c  
LOCUS  
DEFINITION AZ677439 Entamoeba histolytica Sheared DNA Entamoeba histolytica  
genomic, DNA sequence.

ACCESSION AZ677439.1 GI:11814585  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 819)  
Eukaryota; Entamoebidae; Entamoeba.  
Loftus, B., Van Aken, S. and Fraser, C.  
Determination of clone end sequences from Entamoeba histolytica  
HMI:IMSS sheared DNA library  
Unpublished (2000)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjoftus@tigr.org  
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
DNA library  
Seq primer: M13-Forward  
Class: shotgun  
High quality sequence start: 24  
High quality sequence stop: 817.

## FEATURES

source  
1..819  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone.lib="Entamoeba histolytica Sheared DNA"  
/note="Vector: pHOS1; Site\_1: Bst I; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,

C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450. The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).

BASE COUNT 187 a 133 c 82 g 417 t  
ORIGIN

Query Match 6.1%; Score 51.6; DB 17; Length 819;  
Best Local Similarity 45.5%; Pred. No. 0.089; Indels 0; Gaps 0;  
Matches 183; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

QY 91 AGAATAATGATAAACAAGAGGCTTCTACATTAGTGCAGAAAGTCAACATCCCAAGTATATCA 150  
DB 435 AGAATTAGTATATCAAAAACCGGGGAAGCAACAGATAGAAAAGAGAAATTTTAAACA 376

QY 151 CACTTGAAGAAATCTCTGCTGAAGAACTCTATTATTAAGGAACAAATCTCTCACTAAA 210  
DB 375 TTTTATAAATGTTTCTCTAAATTTATTTGAAGAAATTAAGAGGTCACACTAAAGAA 316

QY 211 AAAATTTTCGGACTAAGAAAGATGCTGATATACAAAAGAGAGAGATTTTACAGAGTA 270  
DB 315 GAATTAATAGAAAGTTTATATCATGAATTTGAATTTAAATTAATTAATTAATGAATAT 256

QY 271 GCTCAGGCGATTGATTTTCAAAATTAATTAATCAGGATTTTTCAGGAAGTATTTGGTTAC 330  
DB 255 AAAGAAATTAATAAGAGATATTTAAATTAAGATGAATTTAAAGATAATCAATAGA 196

QY 331 TCTATGACGACCAAGATGAACCTGAGCTGATATCAACATTTAATCCAAAAC 390  
DB 195 ATAATGGAGGAGAAATTAATAAACCAATGGAAGATAAGAAATTTAAAGAAATAATA 136

QY 391 ACCGATACAATGATGATTAATGCTGAATTAATCACTATAACATTTTGCATTATCTGTA 450  
DB 135 TCAAAATGAACAAGAGAGAAATAGAAAGAGAGGATTAATGAATATTTAAAGAA 76

QY 451 GATGCAATGGAAGATCAGCAATATGATGATCTTAAATATGAC 492  
DB 75 GAAGAAATGAAGAAACAAGATGATGATCTTAACTAC 34

RESULT 20  
AZ544101/c  
LOCUS  
DEFINITION  
ENTPF80TF Entamoeba histolytica Sheared DNA linear GSS 14-NOV-2000  
genomic, DNA sequence.

ACCESSION  
AZ544101  
VERSION  
AZ544101.1 GI:11163632  
KEYWORDS  
GSS.  
SOURCE  
Entamoeba histolytica.  
ORGANISM  
Entamoeba histolytica.  
REFERENCE  
1 (bases 1 to 833)  
Eukaryota; Entamoebidae; Entamoeba.  
Loftus, B., Van Aken, S. and Fraser, C.  
Determination of clone end sequences from Entamoeba histolytica  
HMI:IMSS sheared DNA library  
Unpublished (2000)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjlloftus@tigr.org  
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
DNA library  
Seq primer: M13-Forward  
Class: shotgun  
High quality sequence start: 33

High quality sequence stop: 828.

FEATURES  
source

Location/Qualifiers  
1. 833

/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"

/db xref="taxon:5759"

/clone lib="Entamoeba histolytica Sheared DNA"

/note="Vector: pHOS1; Site 1: Bat I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica

using a method described by Clark and Diamond (Clark,

C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a

method for isolate identification. Exp. Parasitol.

77:450.). The DNA was mechanically sheared to give a

tight size distribution (~2 kb). The v + i method used for

the library construction is described in detail in Smith,

H.O. and Venter, J.C. (Making small insert libraries for

whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barrell, Oxford University Press, 1999).

BASE COUNT 196 a 129 c 76 g 432 t

ORIGIN

Query Match 6.1%; Score 51.6; DB 17; Length 833;

Best Local Similarity 45.5%; Pred. No. 0.089;

Matches 183; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

QY 91 AGAATAATGATAAACAAGAGGCTTCTACATTAGTGCAGAAAGTCAACATCCCAAGTATATCA 150

DB 619 AGAATTAGTATATCAAAAACCGGGGAAGCAACAGATAGAAAAGAGAAATTTTAAACA 560

QY 151 CACTTGAAGAAATCTCTGCTGAAGAACTCTATTATTAAGGAACAAATCTCTCACTAAA 210

DB 559 TTTTATAAATGTTTCTCTAAATTTATTTGAAGAAATTAAGAGGTCACACTAAAGAA 500

QY 211 AAAATTTTCGGACTAAGAAAGATGCTGATATACAAAAGAGAGATTTTTCAGAGTA 270

DB 499 GAATTAATAGAAAGTTTATATCATGAATTTGAATTTAAATTAATTAATGAATAT 440

QY 271 GCTCAGGCGATTGATTTTCAAAATTAATTAATCAGGATTTTTCAGGAAGTATTTGGTTAC 330

DB 439 AAAGAAATTAATAAGAGATATTTAAATAAGAAATGAATTTAAAGATAATCAATAGA 380

QY 331 TCTATGACGACCAAGATGAACCTGAGCTGATATCAACATTTTAAATCCAAAAC 390

DB 379 ATAATGGAGGAGAAATTAATAAACCAATGGAAGATAAGAAATTTAAAGAAATAATA 320

QY 391 ACCGATACAATGATGATTAATGCTGAATTAATCACTATAACATTTTGCATTATCTGTA 450

DB 319 TCAAAATGAACAAGAGAGAAATAGAAAGAGAGGATTAATGAATATTTAAAGAA 260

QY 451 GATGCAATGGAAGATCAGCAATATGATGATCTTAAATATGAC 492

DB 259 GAAGAAATGAAGAAACAAGATGATGATCTTAACTAC 218

RESULT 21

AZ545164/c

LOCUS

DEFINITION

ENTW83TR Entamoeba histolytica Sheared DNA linear GSS 14-NOV-2000

genomic, DNA sequence.

ACCESSION

AZ545164

VERSION

AZ545164.1 GI:11165681

KEYWORDS

GSS.

SOURCE

Entamoeba histolytica.

ORGANISM

Entamoeba histolytica.

REFERENCE

1 (bases 1 to 912)

Eukaryota; Entamoebidae; Entamoeba.

AUTHORS

Loftus, B., Van Aken, S. and Fraser, C.

TITLE

Determination of clone end sequences from Entamoeba histolytica

JOURNAL

HMI:IMSS sheared DNA library

Unpublished (2000)

COMMENT

Contact: Brendan J Loftus

```

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlloftus@igr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 56
High quality sequence stop: 900.
FEATURES
    source
        1..912
            /organism="Entamoeba histolytica"
            /strain="HMI:IMSS"
            /db_xref="taxon:5759"
            /clone.lib="Entamoeba histolytica Sheared DNA"
            /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
            Institute for Genomic Research (TIGR), Rockville, MD.
            Genomic DNA isolated from broth cultures of E. histolytica
            using a method described by Clark and Diamond (Clark,
            C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
            method for isolate identification. Exp. Parasitol.
            77:450.). The DNA was mechanically sheared to give a
            tight size distribution (~2 kb). The v + i method used for
            the library construction is described in detail in Smith,
            H.O. and Venter, J.C. (Making small insert libraries for
            whole genome shotgun sequencing projects. In Genome
            Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999)."
```

BASE COUNT 202 a 140 c 102 g 468 t

Query Match 6.1%; Score 51.6; DB 17; Length 912;  
 Best Local Similarity 45.5%; Pred No. 0.09;  
 Matches 183; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

```

Qy 91 AGAACTAATGATAACAAGAAGGCTTCTACATTAGTGCAGAGTACCAATCCAAAGTATATCA 150
Db 493 AGAATTAGATAATCAAAAAACGGGGAGCAACAGATAGAAAAGAGAATTTTAAACA 434
Qy 151 CACTTTAGAAAATTTCTGCTGAGAAACTCTATTATATGGAACAATTTCTCCTCACTAAA 210
Db 433 TTTTATTAATGTTTGTCTTAAATTTATTGAAAGAAAATTAAGAGGTCACACTAAAGAA 374
Qy 211 AAAGTTTTCGACATAAAGAGATGGTGATATACAAAAAGAGAGATTTTACAGAGTA 270
Db 373 GAATAATAGAAAGGTTAATACATGAATTTGAAATTAATAATTAATGATTAATGAATAT 314
Qy 271 GCTCCAGGCATTTGATTTTCAAAATAACTTAATATCAGGATTTTCAGGAAGTATTGGTTAC 330
Db 313 AAAGAAATATAAAGAGATATTTAAATAAAGAAATGAATGAAATTAAGATAATCAATAGA 254
Qy 331 TCTATGGACGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAAAC 390
Db 253 ATAATGGAGGAAGAAATAATAAACCACCAATGGAAGATAAAGAAATTAAGAAAGAAATATA 194
Qy 391 ACCGATACAAATGATACATAATGGTGAATGCTATATCAACATTTTGCATTTATCTCGTAAA 450
Db 193 TCAAAATGAACAAGAAAGAAATGAGAGAGAGGAGGTAATGAATAATTTTAAAGAA 134
Qy 451 GATCGAATGGAGATCAGCAATATGTAGTATCTTAAAAATGAC 492
Db 133 GAAGAAATGAAGACAGAGAGATGTCACAAATTTTAACTAC 92
```

RESULT 22  
 BHI58221/c  
 LOCUS  
 DEFINITION ENT5D60TF Entamoeba histolytica Sheared DNA Entamoeba histolytica  
 genomic, DNA sequence.  
 ACCESSION BHI58221

BHI58221.1 GI:15731659  
 GSS.  
 Entamoeba histolytica.  
 SOURCE  
 Entamoeba histolytica  
 ORGANISM  
 Eukaryota; Entamoebidae; Entamoeba.  
 REFERENCE  
 1 (bases 1 to 912)  
 Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.  
 Determination of clone end sequences from Entamoeba histolytica  
 HMI:IMSS sheared DNA library (2001)  
 Unpublished (2001)  
 JOURNAL  
 COMMENT  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: bjlloftus@igr.org  
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
 DNA library  
 Seq primer: M13-Forward  
 Class: shotgun  
 High quality sequence start: 18  
 High quality sequence stop: 769.  
 FEATURES  
 source  
 1..912
 /organism="Entamoeba histolytica"
 /strain="HMI:IMSS"
 /db\_xref="taxon:5759"
 /clone.lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pHOS1; Site\_1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Barrell, Oxford University Press, 1999)."

BASE COUNT 227 a 137 c 79 g 469 t

Query Match 6.1%; Score 51.6; DB 17; Length 912;  
 Best Local Similarity 45.5%; Pred No. 0.09;  
 Matches 183; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

```

Qy 91 AGAACTAATGATAACAAGAAGGCTTCTACATTAGTGCAGAGTACCAATCCAAAGTATATCA 150
Db 701 AGAATTAGATAATCAAAAAACGGGGAGCAACAGATAGAAAAGAGAATTTTAAACA 642
Qy 151 CACTTTAGAAAATTTCTGCTGAGAAACTCTATTATTAATGGAACAATTTCTCCTCACTAAA 210
Db 641 TTTTATAAATGTTTGTCTTAAATTTATTGAAAGAAAATTAAGAGGTCACACTAAAGAA 582
Qy 211 AAAGTTTTCGACATAAAGAGATGGTGATATACAAAAAGAGAGATTTTACAGAGTA 270
Db 581 GAATAATAGAAAGGTTAATACATGAATTTGAAATTAATAATTAATGATTAATGAATAT 522
Qy 271 GCTCCAGGCATTTGATTTTCAAAATAACTTAATATCAGGATTTTCAGGAAGTATTGGTTAC 330
Db 521 AAAGAAATATAAAGAGATATTTAAATAAAGAAATGAATGAAATTAAGATAATCAATAGA 462
Qy 331 TCTATGGACGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAAAC 390
Db 461 ATATGGAGGAAGAAATAATAAACCACCAATGGAAGATAAAGAAATTAAGAAAGAAATATA 402
Qy 391 ACCGATACAAATGATACATAATGGTGAATGCTATATCAACATTTTGCATTTATCTCGTAAA 450
Db 401 TCAAAATGAACAAGAAAGAAATAGAGAGAGGAGGTAATGAATAATTTTAAAGAA 342
```

[illegible]



```

Db 612 AGAATTAGTAATCAAAAAACGGGGGAAGCAAAACAGATAGAAAAAGAAAATTTTAAACA 553
Qy 151 CACTTTAGAAAATCTCTGCTGAAGAACTCTTATTAATGGAACAAATCTCTCACTAAA 210
Db 552 TTTTATAAATGTTTGGCTCTAAATTAATTTGAAAGAAAATTAAGAGGTCACACTAAAGAA 493
Qy 211 AAAGTTTTCGCACTAAGAAAGATGGTGATATACAAAAGAGAGATTTTACAGAGTA 270
Db 492 GAAATAATAGAAAGGTTAATCATGAATTTGAAATTAATAATTAATTAATTAATGAATAT 433
Qy 271 GCTCAGGCAATGATTTTCAAAATAACCTTAATATCAGGATTTTCAGGAAGTATTGGTTAC 330
Db 432 AAAGAAATAATAAGAGATATTTAAATAAAGATGAATTAAGATAATCAATAGA 373
Qy 331 TCTATGAGCGACCAAGAAATGAACTTGAAGCTGCATATCAACAATTTTAATCCAAAAC 390
Db 372 ATAATGGAGGAAGAAATAATAAACCACTAAGGAAGATAAGAAATTAAGAAATAATA 313
Qy 391 ACCGATAACATGATACCTGATTAATGGTGAATCTATTAACATTTTGCATTTATCTCGTAA 450
Db 312 TCAAAATGAACAAGAAAGAAATAGAAAGAGAGGAGTAAATCAAAATATTTAAAGAA 253
Qy 451 GATGCAATGGAAGATCAGCAATATCTAGTACTTAAATAATGAC 492
Db 252 GAAGAAATGAAGAAACAAGAAATGATCAATTTTAACTAC 211

RESULT 25
LOCUS CNS0187R 878 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN04E04 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL108993
VERSION AL108993.1 GI:5629297
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 878)
AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HM1:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: Shotgun
High quality sequence start: 14
High quality sequence stop: 833.
FEATURES
source
1..884
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pBOS1; Site 1: Bat 1; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT 454 a 67 c 120 g 243 t
ORIGIN
141 AAGTATATCACACTTACAAATTTCTGCTGGAAGAACTCTTATTAATGAAACAATTC 200
Db 43 AAAGAAATCAAAATATGATGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 102

```

```

Qy 201 TCTCACTAAAAAGTTTTCGAGCTAAAGAAAGATGGTGATATACAAAAAAGACGATTT 260
Db 103 MAAAAAAAWMAWARGAWGCGWGGAWMMWMAWRG-MMMWMAWMAAAWMAWMAWCTTT 161
Qy 261 TACAAGAGTAGCTCCAGGCATTGATTTTCAAAATAACTTAATATACAGATTTTTCAGGAG 320
Db 162 GCMCTGATTTGTTTGTWMMRTTAAWMMMAWMAWMAWMAWMAWMAWMAWMAWMAWMAW 221
Qy 321 TATTGGTTACTCTATGACGAGCAAGAAATAGAACTTGAAGCTCATATCAACAATTTAA 380
Db 222 AARAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMA 281
Qy 381 TCCAAAAACACCGATACCAATGATCTGATTAATGGTGAATATCTATATAACATTTTGC 440
Db 282 WMGAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMA 341
Qy 441 ATCTCGTAA 449
Db 342 AARTMRTAW 350

RESULT 26
LOCUS AZ669586 884 bp DNA linear GSS 14-DEC-2000
DEFINITION ENTH07TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION AZ669586
VERSION AZ669586.1 GI:11806732
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 884)
AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HM1:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: Shotgun
High quality sequence start: 14
High quality sequence stop: 833.
FEATURES
source
1..884
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pBOS1; Site 1: Bat 1; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT 454 a 67 c 120 g 243 t
ORIGIN

```

Query Match 6.1%; Score 51.2; DB 17; Length 884;  
Best Local Similarity 45.7%; Pred. No. 0.11;  
Matches 179; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

QY 97 AATGATAACAAGAGGCTTCTACATTAGTGCAGAGTACAATCCCAAGTATATACACACTTT 156  
DB 150 ATTAATAATAAACAATTTGAATAAACAAGATAGAGAGAGAAATTTTACACTTTAT 209  
QY 157 AGAAAAATCTCTGCTGAAGAACTCCTATTAAATGAACAAATCTCTCACTAAAAAGTT 216  
DB 210 AAATGTTTTGCTTTAAAAATGTTTCAAGAAAAATTAAGAGGTCAACTTTAAAGAGAAATA 269  
QY 217 TTCGACTAAGAGAGAGTGCATATACAAAAAGAGCGATTTTACAGAGTAGCTCCA 276  
DB 270 ATAGAAAGGTTAATCAAGAATTTGGAATTTAAATAAATGATTAATTAATGAATATAAGAA 329  
QY 277 GCATTGATTTTCAAAATAAATTAATATCAGGATTTTCAGGAAGTATTGGTTACTCTATG 336  
DB 330 GTAATAAGAGACACTTTTAAATAAAGATGAATTAAGAGAGAAATCAATAGATATATG 389  
QY 337 GAGGACCAAGATAGAACTTGAAGCTGCATATCAACAAATTTAAATCCAAAAACCGAT 396  
DB 390 GAGGAAGAAATAAATAAACAATCAATCGAAAGATAATGAATTAAGAGAAATTAATATCCAAA 449  
QY 397 AACATGATCTACTGATTAATGGTGAATTAATCACTATAACATTTTGCATTATCTCGTAAAGATCA 456  
DB 450 ATGAACAGAAAGAGAAATATAAGAGAGGAGGTAATGAAGTATTTAAGAGAGAGAA 509  
QY 457 ATGGAAGATCAGCAATATGTAGTACTTAAAAA 488  
DB 510 ATGAAGAGACAGAGATGTACAATTTATTAA 541

RESULT 27  
AZ538693/c  
LOCUS ENTGS73TF Entamoeba histolytica Sheared DNA linear GSS 14-NOV-2000  
DEFINITION genomic, DNA sequence.

ACCESSION AZ538693  
VERSION AZ538693.1 GI:11143876  
KEYWORDS GSS.

SOURCE Entamoeba histolytica.  
ORGANISM Entamoeba histolytica  
Eukaryota; Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 898)  
Loftus, B., Van Aken, S. and Fraser, C.

AUTHORS Determination of clone end sequences from Entamoeba histolytica  
TITLE HMI:IMSS sheared DNA library  
JOURNAL Unpublished (2000)

COMMENT Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjloftus@igr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
DNA library

Seq primer: M13-Forward  
Class: shotgun  
High quality sequence start: 28  
High quality sequence stop: 743.

FEATURES  
source

Location/Qualifiers  
1..898

/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"

/db\_xref="taxon:5759"

/clone\_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: pHS1; Site: 1; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a

method for isolate identification. Exp. Parasitol.  
77:450.). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + i method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barrell, Oxford University Press, 1999)."

BASE COUNT 219 a 134 c 69 g 476 t

Query Match 6.1%; Score 51; DB 17; Length 898;

Best Local Similarity 45.9%; Pred. No. 0.12;  
Matches 159; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 154 TTTAGAAAATCTCTGCTGAAGAACTCTATTAAATGGAACAAATCTCTCACTAAAAA 213  
DB 722 TTTATAATGTTGTTTGAATATTTGAAAGAAATTAAGAGGTCAACTTTAAAGAGAG 663  
QY 214 GTTTTCGACTAAGAGAGATGGTGATATACAAAAAGAGCGATTTTACAGAGTAGCT 273  
DB 662 TTAATAGAAAGATTAAATACATGAATTTGGAATTAAGGTAATATATTAATGAATATAA 603  
QY 274 CCAGGCATTGATTTTCAAAATAAATTAATATCAGGATTTTCAGGAAGTATTGGTTACTCT 333  
DB 602 GAATATATAAGAGACATTTAAATTAAGNATGAAATAAAGAGAAATTAATAGATA 543  
QY 334 ATGACGACCAAGATAGAACTTTGAAGCTGCATATCAACAAATTTAAATCCAAAAACACC 393  
DB 542 ATGAGGAAGAAATAATAAACAATCAATGGAAGATAAGAAATTTAAAGAAATTAATACT 483  
QY 394 GATAACAATGATCTGATATGTTGAATGATCTATATAACATTTTGCATTATCTCGTAAAGAT 453  
DB 482 AGAATGAACAAGAAAGAAATAGAAAGAGAGAGGTAATGAAATAATTTTAAAGAGAA 423  
QY 454 GCAATGGAAGATCAGCAATATGTAGTACTTTAAAAATGAC 492  
DB 422 GAATGAAGAGACAGAGAGATGTACAATTTTAACTAC 384

RESULT 28

BH132524

LOCUS

ENTNE36TF Entamoeba histolytica Sheared DNA linear GSS 07-AUG-2001

DEFINITION genomic, DNA sequence.

ACCESSION BH132524

VERSION BH132524.1 GI:15091388

KEYWORDS GSS.

SOURCE Entamoeba histolytica.

ORGANISM Entamoeba histolytica

Eukaryota; Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 932)

Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.

AUTHORS Determination of clone end sequences from Entamoeba histolytica

TITLE HMI:IMSS sheared DNA library (2001)

JOURNAL Unpublished (2001)

COMMENT Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjloftus@igr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared

DNA library

Seq primer: M13-Forward

Class: shotgun

High quality sequence start: 25

High quality sequence stop: 625.

FEATURES

source

Location/Qualifiers  
1..932

/organism="Entamoeba histolytica"

/strain="HMI:IMSS"





```

RESULT 32
AZ529741
LOCUS
DEFINITION
  Entamoeba histolytica Sheared DNA Entamoeba histolytica
  genomic, DNA sequence.
ACCESSION
  AZ529741
KEYWORDS
  GSS.
SOURCE
  Entamoeba histolytica.
  Entamoeba histolytica.
  Eukaryota; Entamoebidae; Entamoeba.
REFERENCE
  1 (bases 1 to 852)
  Loftus, B., Van Aken, S. and Fraser, C.
  Determination of clone end sequences from Entamoeba histolytica
  HMI:IMSS sheared DNA library
  Unpublished (2000)
  Contact: Brendan J Loftus
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0208
  Fax: 301 838 3543
  Email: bjlloftus@tigr.org
  Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
  DNA library
  Seq primer: M13-Reverse
  Class: shotgun
  High quality sequence start: 17
  High quality sequence stop: 852.
  Location/Qualifiers
    1..852
      /organism="Entamoeba histolytica"
      /strain="HMI:IMSS"
      /db_xref="taxon:5759"
      /clone_lib="Entamoeba histolytica Sheared DNA"
      /note="Vector: PHOS1; Site 1: Bst I; Constructed at The
      Institute for Genomic Research (TIGR), Rockville, MD.
      Genomic DNA isolated from broth cultures of E. histolytica
      using a method described by Clark and Diamond (Clark,
      C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
      method for isolate identification. Exp. Parasitol.
      77:450.). The DNA was mechanically sheared to give a
      tight size distribution (~2 kb). The v + i method used for
      the library construction is described in detail in Smith,
      H.O. and Venter, J.C. (Making small insert libraries for
      whole genome shotgun sequencing projects. In Genome
      Sequencing: A Practical Approach, eds M. Vaubin and B.
      Barrell, Oxford University Press, 1999)."
  BASE COUNT      428 a  70 c  125 g      229 t
  ORIGIN
    Query Match      6.0%; Score 50.4; DB 17; Length 852;
    Best Local Similarity 45.6%; Pred. No. 0.17;
    Matches 177; Conservative 0; Mismatches 211; Indels 0; Gaps 0;

  Qy 101 ATACAAAGAGGGCTTACATTAGTGCAGAGTACAAATCCAGTATATACACCTTTAGAA 160
  Db 350 ACAAAAATAAATAATTCGAATTAACACAGAGAGGATAAATTTTAACTTTATATAA 409
  Qy 161 AATTCTCTGCGAAGAACTCCTATTATATGACAAATCTCTCACTAAAAAGTTTCG 220
  Db 410 TGTITGTTGTTGAAATTTTGAAGAAGAAATTAAGATGTCACACTAAAGAGAAATAATAG 469
  Qy 221 GACTAAAGAGAGTGGTGATATAACAAAAAGAGAGATTTTCAAGAGTAGCTCCAGGCA 280
  Db 470 AAAGATTATATATGAATTCGAATTAATAATTAATGATATTATGAATATAAGAGATTA 529
  Qy 281 TTGATTTTCAAAATAACTTAATATACAGATTTTCAGGAAGTATTGGTTACTCTATGGACG 340
  Db 530 TAAAGAAGATATTTAAAAATAAAGATGAATGAATTTAAAGAGAAATAAATAGAGAGATGGAGG 589
  Qy 341 GACCAGATAGAACTTGAAGCTGCATATCAACAAATTTAATCCAAAAAACCCGATACCA 400

```

---

```

Db 590 AAGAAATAATAAATGTCATAGAGAAATAGTAATGAAATTAAGAAATTAATACTAGAAATGA 649
Qy 401 ATGATCTAGTAATAGGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 460
Db 650 ATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 709
Qy 461 AAGATCAGCAATATGTAGTACTTTAAAAA 488
Db 710 AAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 737

RESULT 33
BI323673
LOCUS
DEFINITION
  Strongyloides ratti L2 pAMP1 v1 Chiapelli McCarter
  Strongyloides ratti cDNA 5' similar to TR:Q94155 Q94155 UNC-40 ;
  mRNA sequence.
ACCESSION
  BI323673
VERSION
  BI323673.1 GI:15002859
KEYWORDS
  EST.
SOURCE
  Strongyloides ratti.
  Strongyloides ratti.
  Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
  Panagrolaimoidea; Strongyloidea; Strongyloides.
REFERENCE
  1 (bases 1 to 481)
  McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,
  Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,
  Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagaris, V., R.,
  Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe
  , M., Allen, M., Person, S., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,
  Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and
  Wilson, R.
  The Washington Univ. Nematode EST Project, 1999
  Unpublished (1999)
  Contact: McCarter JP
  The Washington Univ. Nematode EST Project, 1999
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: est@watson.wustl.edu
  The library was constructed by Brandi Chiapelli and Dr. James
  McCarter (bchiapelli@watson.wustl.edu & jmcarter@watson.wustl.edu) at
  Washington University, St. Louis. DNA Sequencing by: Washington
  University Genome Sequencing Center St. Louis.
  Seq primer: -40SP from Gibco
  High quality sequence stop: 389.
  Location/Qualifiers
    1..481
      /organism="Strongyloides ratti"
      /db_xref="taxon:34506"
      /clone_lib="Strongyloides ratti L2 pAMP1 v1 Chiapelli
      McCarter"
      /dev_stage="L2"
      /lab_host="DH10B"
      /note="vector: pAMP1 (Gibco); The library was constructed
      by Brandi Chiapelli and Dr. James McCarter at Washington
      University, St. Louis. The cDNA was made by using
      Dynabead oligo-dT priming (Dynal). PCR based library
      using a modified protocol from the SMART PCR cDNA
      Synthesis Kit from Clontech. Directionally cloned into the
      UDG sites of pAMP1. Nematodes were provided by Dr. Mark
      Viney of Bristol, UK."
  BASE COUNT      186 a  86 c  58 g      151 t
  ORIGIN
    Query Match      6.0%; Score 50; DB 13; Length 481;
    Best Local Similarity 50.9%; Pred. No. 0.2;
    Matches 119; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

  Qy 241 ATACAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
  Db 128 ACATCTACCACTGCTCTCTCAATGGATCTCCAGGTATTATTAAACCACTACATTTCA 187

```

QY 301 ATATCAGGATTTTCAGGAGTATTGGTCTACTCTATGAGCGGACCAAGAAATAGAACTTGA 360  
 Db 188 CATTCTCAGACATCAGTCTCTATTTGGTAGAGCTTCTGCACACCAAGAGTAATATGCGCA 247  
 QY 361 GCTGCATATCAACAATTTAATCCAAAAACACCGATAACAATGATGATAATGGTGA 420  
 Db 248 AGTATATATACACGAGTGTAGTCCATATAAAAAATATCAAAATATGTTATTGGATAAAT 307  
 QY 421 TACTATAACATTTTTCGATTTCTCGTAAAGATGCAATGGAAGATCAGCAATPAT 474  
 Db 308 GATAATGAACAGATTCCAGATATAGTACTACTGCTGGTGGTTTAAATCAAGAT 361

RESULT 34  
 BH167558  
 LOCUS  
 DEFINITION ENTRZ50TR Entamoeba histolytica Sheared DNA linear GSS 24-SEP-2001  
 genomic, DNA sequence.

ACCESSION  
 VERSION BH167558.1 GI:15740996  
 KEYWORDS  
 SOURCE Entamoeba histolytica.

ORGANISM  
 Entamoeba histolytica  
 Eukaryota; Entamoebidae; Entamoeba.

REFERENCE  
 1 (bases 1 to 885)  
 Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.

AUTHORS  
 TITLE Determination of clone end sequences from Entamoeba histolytica  
 HMI:IMSS sheared DNA library (2001)

JOURNAL  
 COMMENT Unpublished (2001)

CONTACT: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: bjloftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
 DNA library

Seq primer: M13-Reverse  
 Class: shotgun

High quality sequence start: 70  
 High quality sequence stop: 534.

## FEATURES

source

1..885

/organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db\_xref="taxon:5759"

/clone\_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: pHOS1; Site 1: Bst I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica

using a method described by Clark and Diamond (Clark,

C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a

method for isolate identification. Exp. Parasitol.

77:450.). The DNA was mechanically sheared to give a

tight size distribution (~2 kb). The v + i method used for

the library construction is described in detail in Smith,

H.O. and Venter, J.C. (Making small insert libraries for

whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barell, Oxford University Press, 1999)."

## BASE COUNT

444 a 79 c 140 g 222 t

## ORIGIN

Query Match

Best Local Similarity 6.0%; Score 50; DB 17; Length 885;

Matches 182; Conservative 0; Mismatches 220; Indels 0; Gaps 0;

## QY

91 AGAATATGATACAAAGAGGCTTCTCATTTAGTGGCAAGTCAATCCAAAGTATATCA 150

## Db

157 AGAATATGATATCAAAAAACGGGGAGCAACAGATAGAAAAAGAAATTTTAAACA 216

QY 151 CACTTAGAAAATCTCTGCTGAGAAACCTCTATTATGGAACAAATCTCTCACTAAA 210  
 Db 217 TTTTATAAATGTTTGTCTTAAATATTTTGAAGAAAATTAAGAGGTCAACTAAAGAA 276  
 QY 211 AAAGTTTTTCGACTAAAGAGAGATGGTGATATAACAAAAAGACGATTTTACAGAGTA 270  
 Db 277 GAATAATAGAAAGGTTAATACATGAATTTGAAATTTAAATATAATGATATTAATGAATAT 336  
 QY 271 GCTCAGGATGATTTTCAAAATACTTAATATCAGGATTTTCAGGAAGTATTTGGTTAC 330  
 Db 337 AAAGAAATATAAAGAGAGATATTTAAATAAAGAAATGAAATTTAAAGATAATCAATAGA 396  
 QY 331 TCTATGACGCGCAAGAAATAGAACTTCAAGCTCATATCAACAATTTAAATCCAAAAAAC 390  
 Db 397 ATAATGGAGGAAGAAATAATAAACCACCAATGGAAGATAAAGAAATTAAGAATAATA 456  
 QY 391 ACCGATACCAATGATCTGATAAATGGTCAATCTATATAACATTTTGCATTATCTCGTAAA 450  
 Db 457 TCAAAATGACGAGAGAAAGAAATAGAGAAGAGGGTAAATGAAATATTTAAAGAA 516  
 QY 451 GATCAATGGAAGATCAGCAATATGTAGTACTTAAAAATGAC 492  
 Db 517 GAAGAAATGAGAGAACCAAGAGATGTACAATTTTAACTAC 558

## RESULT 35

BH146655

LOCUS

DEFINITION ENTPB52TF Entamoeba histolytica Sheared DNA linear GSS 27-AUG-2001

genomic, DNA sequence.

ACCESSION

VERSION BH146655

KEYWORDS

SOURCE

ORGANISM

Entamoeba histolytica.

Eukaryota; Entamoebidae; Entamoeba.

REFERENCE

AUTHORS

LOftus,B., Wang,Z., Van Aken,S. and Fraser,C.

TITLE Determination of clone end sequences from Entamoeba histolytica

HMI:IMSS sheared DNA library (2001)

JOURNAL

COMMENT Unpublished (2001)

CONTACT: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjloftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared

DNA library

Seq primer: M13-Forward

Class: shotgun

High quality sequence start: 34

High quality sequence stop: 817.

Location/Qualifiers

1..886

/organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db\_xref="taxon:5759"

/clone\_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: pHOS1; Site 1: Bst I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica

using a method described by Clark and Diamond (Clark,

C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a

method for isolate identification. Exp. Parasitol.

77:450.). The DNA was mechanically sheared to give a

tight size distribution (~2 kb). The v + i method used for

the library construction is described in detail in Smith,

H.O. and Venter, J.C. (Making small insert libraries for

whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barell, Oxford University Press, 1999)."

BASE COUNT 453 a 76 c 131 g 226 t

Query Match 6.0%; Score 50; DB 17; Length 886;  
Best Local Similarity 48.3%; Pred. No. 0.21;  
Matches 140; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 202 CTCACATAAAGAGTTTCGGACTAAGAAAGATGCTGATATACAAAAGACGATTTT 261  
DB 322 CTTAAGAAGAATAATAGAAGATTAATACATGAATTTGGAAATTTAAAGTTAATGATATT 381

QY 262 ACAAGTAGCTCCAGGCAATGATTTTCAAAATAACTTAATATCAGGATTTTCAGGAAGT 321  
DB 382 AATGAATATAAGAAATTAATAAGAGACATTTAAATAAAGATGAATTAAGAGAA 441

QY 322 ATTGTTACTCTATGAGCGGACCAAGATAGAACTTTGAAGCTGCAATCAACAATTTAAT 381  
DB 442 ATCAATAGATAATGAGGAGAAATATAAACAATCAATGTCAGATTAATGATTAATA 501

QY 382 CCAAAAACCGGATAACAATGATCTGATTAATGCTGATTAATCAATTAACATTTGCAATTA 441  
DB 502 GAAATAATACTAGAAATCAACAAGAAAGAAATAGAAAGATTTGGAAGTAAATGAAATA 561

QY 442 TCTCGTAAAGATCAATGCAAGATCAGCAATATGATGACTTAAATAATGA 491  
DB 562 TTTAAGAAGAATAATGAAGAACAAGAGATGTACAAATTTATTAAATTA 611

RESULT 36  
BH152307/c

LOCUS ENTPO68TR Entamoeba histolytica Sheared DNA Entamoeba histolytica  
DEFINITION genomic, DNA sequence.

ACCESSION BH152307  
VERSION BH152307.1 GI:15315989  
KEYWORDS GSS  
SOURCE Entamoeba histolytica.  
ORGANISM Eukaryota; Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 908)  
Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.  
Determination of clone end sequences from Entamoeba histolytica  
HMI:IMSS sheared DNA library (2001)  
Unpublished (2001)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjloftus@tigr.org  
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library  
Seq primer: M13-Reverse  
Class: shotgun  
High quality sequence start: 17  
High quality sequence stop: 699.

FEATURES  
source  
1..908  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"  
/note="Vector: pHO1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for

whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999).

BASE COUNT 225 a 145 c 90 g 448 t

Query Match 6.0%; Score 50; DB 17; Length 908;  
Best Local Similarity 45.3%; Pred. No. 0.21;  
Matches 182; Conservative 0; Mismatches 220; Indels 0; Gaps 0;

QY 91 AGAATTAATGATAACAAGAGGCTTCTACATTAGTCARAGTCAATCCCAAGTATATCA 150  
DB 551 AGAATTAAGACAGCAAAAACGGGGAGCAACAGATAGAAAAGAGAAATTTTAAACA 492

QY 151 CACTTTAGAAAATCTCTGCTGAAGAAACTCTTAAATGGAACAAATTCCTCACTAAA 210  
DB 491 CTTTATAAATGTTTTTTTGAATTTTGAAGAAAATTTAAGAGGTCACTTAAAGAA 432

QY 211 AAAGTTTCGGACATAAAGAAAGATGGTGATATACAAAAGAGAGATTTTACAGAGTA 270  
DB 431 GAGTTAATAGAAAGATTAATACATGAATTTTGAATTTAAAGTAAATTAATTAATGAATAT 372

QY 271 GCTCCAGGCATTGATTTTCAAAATAACTTAAATATCAGGATTTTCAGGAAGTATTGGTTAC 330  
DB 371 AAAGAAATATAAAGAGACATTTTAAATAAAGATGAATAAAGAGAAATTAATAGA 312

QY 331 TCTATGCGGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTTAAATCCAAAAC 390  
DB 311 ATAATGAGGAGAAATATAAACAATCAATCGAAGATTAAGATTTAAAGAAATATA 252

QY 391 ACCGATTAACATGATCTGATAATGTTGATATATAACATTTTGCATTTCTCGTAAA 450  
DB 251 ACTAGAATGAACAAGAAAGAAATAGAAAGAGGAGGTAATGAAATAATTTTAAAGAA 192

QY 451 GATCCAATGGAGATCAGCATATGATAGTACTTAAATAATGAC 492  
DB 191 GAAGAAATGAAGAAACAAGAGATGTACAAATTTTATTAACTAC 150

BH160886 936 bp DNA linear GSS 24-SEP-2001  
ENTRO55TR Entamoeba histolytica Sheared DNA Entamoeba histolytica  
DEFINITION genomic, DNA sequence.

ACCESSION BH160886  
VERSION BH160886.1 GI:15734324  
KEYWORDS GSS  
SOURCE Entamoeba histolytica.  
ORGANISM Eukaryota; Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 936)  
Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.  
Determination of clone end sequences from Entamoeba histolytica  
HMI:IMSS sheared DNA library (2001)  
Unpublished (2001)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjloftus@tigr.org  
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library  
Seq primer: M13-Reverse  
Class: shotgun  
High quality sequence start: 21  
High quality sequence stop: 786.

FEATURES  
source  
1..936  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"

/clone.lib="Entamoeba histolytica Sheared DNA"  
 /note="Vector: pHS1; Site: Bst I; Constructed at The  
 Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from broth cultures of E. histolytica  
 using a method described by Clark and Diamond (Clark,  
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
 method for isolate identification. Exp. Parasitol.  
 77:450.). The DNA was mechanically sheared to give a  
 tight size distribution (~2 kb). The v + i method used for  
 the library construction is described in detail in Smith,  
 H.O. and Venter, J.C. (Making small insert libraries for  
 whole genome shotgun sequencing projects. In Genome  
 Sequencing: A Practical Approach, eds M. Vaudin and B.  
 Barrell, Oxford University Press, 1999)."

BASE COUNT 470 a 78 c 131 g 257 t  
 ORIGIN  
 Query Match 6.0%; Score 50; DB 17; Length 936;  
 Best Local Similarity 48.3%; Pred. No. 0.21;  
 Matches 140; Conservative 0; Mismatches 150; Indels 0; Gaps 0;  
 QY 202 CTCACATAAAAGTTTCGGACTAAAGAAAGATGGTGATATAACAAAAAGACGATTTT 261  
 DB 134 CTTAAAGGAAGAAATAATAGAAAGATTATACATGAATTTGGAAATTAAAGTTAATGATTT 193  
 QY 262 ACAAGAGTAGCTCCAGGATTCATTTTCAAAATACCTTAATATCAGGATTTTCAGGAAGT 321  
 DB 194 AATGAATATAAAGAAATATAAAGAGACATTTTAAATTAAGAAATGAATTAAGAGAAA 253  
 QY 322 ATTGGTTACTCTATGACGGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAAT 381  
 DB 254 ATCAATAGATTAATGGAGAGAAATATAAATCAATCATGTCAGATAATGCAATTAATA 313  
 QY 382 CCAAAAAACACGATAACATGATCTGATATGGTGATGATATACTATAACATTTTCATTA 441  
 DB 314 GAAATAATAACTAGAAATGAACAAAGAAAGAAATAGAAAGATTTGGAAGTAAATGAATA 373  
 QY 442 TCTCGTAAGATGCATGCAAGATCAGCAATATGTAGTACTTTAAATGA 491  
 DB 374 TTAAAGGAAGAAATGAAGAAACAAAGAAAGATGTACAAATTTATTAAATTA 423

RESULT 38  
 CNS0167M/c 1201 bp DNA linear GSS 26-JUL-1999  
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC  
 DEFINITION BACN15M24 of DrosBAC library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.  
 ACCESSION AL106396  
 VERSION AL106396.1 GI:5621701  
 KEYWORDS GSS.  
 SOURCE Drosophila melanogaster.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 1201)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 COMMENT Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the European Drosophila Genome Project (EDGP) -  
 http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC  
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
 d'Etude du Polymorphisme Humain) with funding provided by a MRC  
 project grant. The DNA was prepared from embryos by Alain Bucheton  
 and Genevieve Payan. It has been constructed in the vector  
 pBelBAC11.  
 FEATURES  
 source Location/Qualifiers  
 1..1201  
 /organism="Drosophila melanogaster"

/db.xref="taxon:7227"  
 /clone="BACN15M24"  
 /clone.lib="DrosBAC"  
 /plasmid="pBelBAC11"  
 /note="end : T7"  
 BASE COUNT 323 a 87 c 79 g 551 t 161 others  
 ORIGIN  
 Query Match 6.0%; Score 50; DB 17; Length 1201;  
 Best Local Similarity 38.3%; Pred. No. 0.21;  
 Matches 152; Conservative 35; Mismatches 210; Indels 0; Gaps 0;  
 QY 90 AAGAACTAATCATACAAAGAGCGCTTCTACATTAGTCRAAGTACAAATCCAAAGTATATC 149  
 DB 1148 AAATATATAAAAAAATAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1089  
 QY 150 ACACCTTTAGAAAATTTCTCTGCTGAAGAACTCTTATTAAATGGAACAAATTTCTCTCACTAA 209  
 DB 1088 AAAAAAATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1029  
 QY 210 AAAAGTTTTCGGACTAAAGAAAGATGGTGATATAACAAAAAGACGATTTTACAGAGT 269  
 DB 1028 TTTTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 969  
 QY 270 AGCTCCAGGCAATTCATTTTCAAAATAACTTAATATCAGGATTTTTCAGGAAGTATTGGTTA 329  
 DB 968 TTTTATATATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 909  
 QY 330 CTCTATGACGCGACCAAGAAATAGAACTTGAAGCTGCATATCAACAATTTTATCCAAAAA 389  
 DB 908 TTTTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 849  
 QY 390 CACGATATAACATGATCTGATTAATGGTGATGATCTATAACATTTTGCATATATCTCGTAA 449  
 DB 848 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 789  
 QY 450 AGATGCAATGCAAGTACAGCAATATGTAGTACTTTAA 486  
 DB 788 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 752

RESULT 39  
 BI815836 560 bp mRNA linear EST 03-OCT-2001  
 LOCUS PfEST0a33c07.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium  
 DEFINITION falciparum cDNA 5', mRNA sequence.  
 ACCESSION BI815836  
 VERSION BI815836.1 GI:15909489  
 KEYWORDS EST.  
 SOURCE malaria parasite P. falciparum.  
 ORGANISM Plasmodium falciparum  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 REFERENCE 1 (bases 1 to 560)  
 AUTHORS Tang, K., Cole, R., Chakrabarti, D., Hayward, R., Clifton, S., Pape, D.,  
 Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B.,  
 Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I.,  
 Tsagaris, R., Belaygorod, L., Franklin, C., Carr, L., Crow, A.,  
 Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D.,  
 Waterston, R., Wilson, R. and Sibley, D.  
 WASHU Plasmodium EST Project  
 UNPUBLISHED (2001)  
 CONTACT: L. David Sibley  
 WASHU Plasmodium EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 Library was constructed by Debopam Chakrabarti DNA sequencing by:  
 Washington University Genome Sequencing Center for information on  
 obtaining a clone please contact: L. David Sibley  
 (sibley@borcim.wustl.edu), Washington University  
 Seq primer: -40UP from Gibco



Search completed: July 4, 2003, 01:02:01  
Job time : 1093 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: July 4, 2003, 01:05:07 ; Search time 1654 Seconds  
(without alignments)  
14780.143 Million cell updates/sec  
Title: US-10-062-624-41  
Perfect score: 840  
Sequence: 1 atgaattataagaataattct.....ttggaatgaggttcacattc 840

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0  
Searched: 2054640 seqs, 14551402878 residues  
Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 9  
Maximum DB seq length: 2000000000  
Post-processing: Listing first 1000 summaries

Database : GenEmbl.\*

- 1: gb ba.\*
- 2: gb hg.\*
- 3: gb in.\*
- 4: gb om.\*
- 5: gb ov.\*
- 6: gb pat.\*
- 7: gb ph.\*
- 8: gb pl.\*
- 9: gb pr.\*
- 10: gb ro.\*
- 11: gb sts.\*
- 12: gb sy.\*
- 13: gb un.\*
- 14: gb vi.\*
- 15: em ba.\*
- 16: em fun.\*
- 17: em hum.\*
- 18: em in.\*
- 19: em nu.\*
- 20: em om.\*
- 21: em or.\*
- 22: em ov.\*
- 23: em pat.\*
- 24: em ph.\*
- 25: em pl.\*
- 26: em ro.\*
- 27: em sts.\*
- 28: em un.\*
- 29: em vi.\*
- 30: em htg hum.\*
- 31: em htg inv.\*
- 32: em htg other.\*
- 33: em htg mus.\*
- 34: em htg pln.\*
- 35: em htg rod.\*
- 36: em htg mam.\*
- 37: em htg vrt.\*
- 38: em ey.\*
- 39: em htgo hum.\*
- 40: em htgo mus.\*
- 41: em htgo other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	840	100.0	6913	1	AF324792 Ehrlichia
2	840	100.0	11329	1	AF082744 Ehrlichia
3	840	100.0	28254	1	AF078553 Ehrlichia
4	30	3.6	27190	1	U72291 Ehrlichia C
5	23	2.7	2362	1	AF319940 Cowdria r
6	23	2.7	3507	1	AF125276 Cowdria r
7	23	2.7	3535	1	AF125274 Cowdria r
8	23	2.7	3538	1	AF125277 Cowdria r
9	23	2.7	3541	1	AF125279 Cowdria r
10	23	2.7	3551	1	AF125275 Cowdria r
11	23	2.7	3572	1	AF125278 Cowdria r
12	23	2.7	63124	2	AC100255 Mus muscu
13	22	2.6	182917	9	AL139383 Human DNA
14	22	2.6	206746	2	AC124404 Mus muscu
15	21	2.5	73435	2	AC094268 Rattus no
16	21	2.5	100927	9	AL157828 Human DNA
17	21	2.5	108867	10	AL732388 Mouse DNA
18	21	2.5	146677	2	AF004725 Oryza sat
19	21	2.5	148919	2	AC020680 Homo sapi
20	21	2.5	151865	2	AC025197 Homo sapi
21	21	2.5	169172	2	AC112846 Rattus no
22	21	2.5	189969	2	AC113984 Mus muscu
23	21	2.5	180931	2	AC115013 Mus muscu
24	21	2.5	181405	9	AC018682 Homo sapi
25	21	2.5	199116	9	AF000871 Homo sapi
26	21	2.5	201306	2	AC102104 Mus muscu
27	21	2.5	211344	9	AP002453 Homo sapi
28	20	2.4	14679	9	AC125236 Homo sapi
29	20	2.4	40629	2	AC129525 Mus muscu
30	20	2.4	60058	2	AC115981 Mus muscu
31	20	2.4	100921	9	AC004996 Homo sapi
32	20	2.4	106253	8	AP003815 Oryza sat
33	20	2.4	139629	2	AP003809 Oryza sat
34	20	2.4	139848	2	AP004335 Oryza sat
35	20	2.4	146805	2	AL450424 Homo sapi
36	20	2.4	153023	2	AC021462 Homo sapi
37	20	2.4	165739	9	AL353728 Human DNA
38	20	2.4	166233	2	AC048369 Homo sapi
39	20	2.4	167076	9	AL606489 Human DNA
40	20	2.4	177032	9	CNS06C7W Human chr
41	20	2.4	180768	2	AC121058 Rattus no
42	20	2.4	185002	2	AC107760 Mus muscu
43	20	2.4	198729	2	AC113292 Mus muscu
44	20	2.4	198788	8	ATCHRIV46 Arabidops
45	20	2.4	201768	2	AC126676 Mus muscu
46	20	2.4	207243	2	AC122908 Mus muscu
47	20	2.4	207674	8	ATFCAB Arabidops
48	19	2.3	505	1	AF165815 Ehrlichia
49	19	2.3	505	1	AF287963 Ehrlichia
50	19	2.3	505	1	AF287964 Ehrlichia
51	19	2.3	505	1	AF287964 Ehrlichia
52	19	2.3	828	1	AF368007 Cowdria r
53	19	2.3	828	1	AF368014 Cowdria r
54	19	2.3	1564	1	CRU49843 Cowdria rum
55	19	2.3	1973	9	BC016872 Homo sapi
56	19	2.3	2228	9	AK094672 Homo sapi
57	19	2.3	2850	10	AB047843 Rattus no
58	19	2.3	5375	8	AF104924 Zea mays
59	19	2.3	5979	6	AX277855 Sequence
60	19	2.3	5979	6	AX235338 Sequence
61	19	2.3	6179	6	AX251097 Sequence
62	19	2.3	6179	6	AX344246 Sequence
63	19	2.3	7243	9	HSM803449 Homo sapi
64	19	2.3	14759	1	AF230642 Ehrlichia
65	19	2.3	27613	3	CEK08G2 Caenorhabd

c 66	19	2.3	41718	3	AF077537	AF077537 Caenorhab	139	19	2.3	172525	9	AC008514	Homo sapi
c 67	19	2.3	46849	9	AC113358	AC113358 Homo sapi	140	19	2.3	173828	2	AC120647	Rattus no
c 68	19	2.3	49091	9	AL390794	AL390794 Human DNA	c 141	19	2.3	175752	2	AC091721	Pan trogi
c 69	19	2.3	49144	2	AC113380	AC113380 Strongylo	c 142	19	2.3	176226	2	AC097368	Homo sapi
c 70	19	2.3	53944	2	HS50662A	Z82901 Human DNA s	c 143	19	2.3	178152	9	AC096633	Homo sapi
c 71	19	2.3	58784	2	AC100989	AC100989 Mus muscu	c 144	19	2.3	178443	2	AC123892	Rattus no
c 72	19	2.3	69552	9	AL365275	AL365275 Human DNA	c 145	19	2.3	178887	9	AC093311	Homo sapi
c 73	19	2.3	74319	9	AL391622	AL391622 Human DNA	c 146	19	2.3	179447	2	AC060806	Homo sapi
c 74	19	2.3	76052	2	AC023262	AC023262 Homo sapi	c 147	19	2.3	179678	2	AC119585	Rattus no
c 75	19	2.3	76488	2	AC090251	AC090251 Homo sapi	c 148	19	2.3	180253	2	AL513362	Homo sapi
c 76	19	2.3	76900	2	AC060787	AC060787 Homo sapi	c 149	19	2.3	182268	2	AC110036	Mus muscu
c 77	19	2.3	80723	9	AC113348	AC113348 Homo sapi	c 150	19	2.3	182860	9	AC068700	Homo sapi
c 78	19	2.3	81924	10	AL731562	AL731562 Mouse DNA	c 151	19	2.3	184368	2	AC113518	Mus muscu
c 79	19	2.3	87695	8	AP004579	AP004579 Lotus jap	c 152	19	2.3	185357	2	AC118156	Rattus no
c 80	19	2.3	87869	9	AC004992	AC004992 Homo sapi	c 153	19	2.3	187235	2	AC113428	Homo sapi
c 81	19	2.3	97693	9	AC002383	AC002383 Homo sapi	c 154	19	2.3	187605	2	AC126588	Rattus no
c 82	19	2.3	100000	9	AB020867	AB020867 Homo sapi	c 155	19	2.3	187778	2	AC034166	Homo sapi
c 83	19	2.3	103125	8	AC005560	AC005560 Arabidops	c 156	19	2.3	188324	2	AC087625	Homo sapi
c 84	19	2.3	105079	9	AL513308	AL513308 Human DNA	c 157	19	2.3	188682	2	AC124825	Mus muscu
c 85	19	2.3	105604	9	AL584013	AL584013 Homo sapi	c 158	19	2.3	188741	2	AC113966	Mus muscu
c 86	19	2.3	107418	9	AL590004	AL590004 Human DNA	c 159	19	2.3	190478	2	AC130127	Rattus no
c 87	19	2.3	109097	2	AC013377	AC013377 Homo sapi	c 160	19	2.3	190706	2	AC009485	Homo sapi
c 88	19	2.3	110000	2	AC107310	AC107310 Homo sapi	c 161	19	2.3	191219	2	AC105508	Rattus no
c 89	19	2.3	110294	2	AC112567	AC112567 Rattus no	c 162	19	2.3	192524	2	AC120742	Rattus no
c 90	19	2.3	113006	9	AC006978	AC006978 Homo sapi	c 163	19	2.3	194612	9	AC021118	Homo sapi
c 91	19	2.3	114100	9	AL583806	AL583806 Human DNA	c 164	19	2.3	194763	2	AC026740	Homo sapi
c 92	19	2.3	118327	8	AC004521	AC004521 Arabidops	c 165	19	2.3	197780	2	AC102754	Mus muscu
c 93	19	2.3	123280	2	AC111706	AC111706 Dictyoste	c 166	19	2.3	199558	2	AC107309	Homo sapi
c 94	19	2.3	124704	2	AL837520	AL837520 Mus muscu	c 167	19	2.3	201140	2	AC016942	Homo sapi
c 95	19	2.3	129677	2	AC113585	AC113585 Trypanoso	c 168	19	2.3	204570	2	AC113498	Mus muscu
c 96	19	2.3	133181	9	AC100854	AC100854 Homo sapi	c 169	19	2.3	206066	2	AC117194	Mus muscu
c 97	19	2.3	133343	2	AC097814	AC097814 Rattus no	c 170	19	2.3	207691	9	AC093917	Homo sapi
c 98	19	2.3	133589	8	AP004576	AP004576 Lotus jap	c 171	19	2.3	209216	9	AF190464	Homo sapi
c 99	19	2.3	133522	9	AC092831	AC092831 Homo sapi	c 172	19	2.3	209802	2	AL772387	Mus muscu
c 100	19	2.3	135444	2	AC106696	AC106696 Rattus no	c 173	19	2.3	223204	2	AC121601	Mus muscu
c 101	19	2.3	142656	9	AC012049	AC012049 Homo sapi	c 174	19	2.3	223624	2	AC093626	Homo sapi
c 102	19	2.3	143792	2	AC125868	AC125868 Rattus no	c 175	19	2.3	224075	2	AC068250	Homo sapi
c 103	19	2.3	145254	9	AP002372	AP002372 Homo sapi	c 176	19	2.3	231177	2	AC095564	Rattus no
c 104	19	2.3	145435	9	AC011419	AC011419 Homo sapi	c 177	19	2.3	232228	2	AC127332	Mus muscu
c 105	19	2.3	145980	9	HS302D9	Z82198 Human DNA s	c 178	19	2.3	236281	9	AC004673	Homo sapi
c 106	19	2.3	146678	2	AC106423	AC106423 Rattus no	c 179	19	2.3	239340	1	AP000993	Thermopla
c 107	19	2.3	148651	2	AC060778	AC060778 Homo sapi	c 180	19	2.3	244106	2	AC095681	Rattus no
c 108	19	2.3	149333	2	AP004592	AP004592 Oryza sat	c 181	19	2.3	340000	9	HS21C009	Homo sapi
c 109	19	2.3	149726	2	AC016394	AC016394 Homo sapi	c 182	19	2.3	300	6	AX389636	Sequence
c 110	19	2.3	150152	2	AL355502	AL355502 Homo sapi	c 183	18	2.1	416	10	AF016505	Rattus no
c 111	19	2.3	150535	2	AC121470	AC121470 Rattus no	c 184	18	2.1	418	11	G61474	G61474 SHGC-85670
c 112	19	2.3	150572	2	CNS0181C	AL163151 Human chr	c 185	18	2.1	586	1	AF095596	Staphyloc
c 113	19	2.3	150635	9	AL365185	AL365185 Human DNA	c 186	18	2.1	800	6	AF032211	Sequence
c 114	19	2.3	151040	2	AC068389	AC068389 Homo sapi	c 187	18	2.1	840	1	AF368009	Cowdria r
c 115	19	2.3	152395	9	AC123022	AC123022 Homo sapi	c 188	18	2.1	840	1	AF368012	Cowdria r
c 116	19	2.3	153387	30	AC027432	AC027432 Homo sapi	c 189	18	2.1	956	6	AR003635	Sequence
c 117	19	2.3	154245	10	AL805895	AL805895 Mouse DNA	c 190	18	2.1	956	8	ATU58919	US8919 Arabidops
c 118	19	2.3	154939	2	AC022537	AC022537 Homo sapi	c 191	18	2.1	964	8	AF113923	Arabidops
c 119	19	2.3	157241	2	AC046194	AC046194 Homo sapi	c 192	18	2.1	1087	8	AF125576	Arabidops
c 120	19	2.3	157875	2	AC007833	AC007833 Homo sapi	c 193	18	2.1	1156	8	AF370220	Arabidops
c 121	19	2.3	157935	2	AC129303	AC129303 Homo sapi	c 194	18	2.1	1204	5	GSU31987	U31987 Gallus gall
c 122	19	2.3	159311	2	HS279F22	AL862879 Homo sapi	c 195	18	2.1	1310	3	EDU85823	US823 Entamoeba d
c 123	19	2.3	160865	2	AC009818	AC009818 Homo sapi	c 196	18	2.1	1417	9	F229081509	AF229089 Homo sapi
c 124	19	2.3	162930	2	AC107706	AC107706 Mus muscu	c 197	18	2.1	1756	9	HSBARD15	AF038038 Homo sapi
c 125	19	2.3	165783	2	AC106579	AC106579 Rattus no	c 198	18	2.1	1904	5	MGU92082	U92082 Meleagris g
c 126	19	2.3	166382	2	AC097341	AC097341 Rattus no	c 199	18	2.1	1990	8	SCAR03	X13514 Yeast AR03
c 127	19	2.3	167133	2	AC016986	AC016986 Homo sapi	c 200	18	2.1	2294	9	HSMATAB	XS5503 H sapiens p
c 128	19	2.3	167758	2	AC091374	AC091374 Rattus no	c 201	18	2.1	2344	5	AB005283	AB005283 Callus ga
c 129	19	2.3	167758	2	AC110189	AC110189 Mus muscu	c 202	18	2.1	2356	14	TOPEBR2	X15883 Fea early b
c 130	19	2.3	167818	9	AC093871	AC093871 Homo sapi	c 203	18	2.1	3248	8	AF125577	AF125577 Arabidops
c 131	19	2.3	168026	9	AL445670	AL445670 Human DNA	c 204	18	2.1	3374	14	PEBYCOPR	X78455 Fea early b
c 132	19	2.3	168849	2	AC105336	AC105336 Mus muscu	c 205	18	2.1	3374	14	TOPEBRN2	X51828 Fea early b
c 133	19	2.3	168907	2	AC023138	AC023138 Homo sapi	c 206	18	2.1	3502	3	DMU27181	U27181 Drosophila
c 134	19	2.3	169794	2	AC004688	AC004688 Plasmodiu	c 207	18	2.1	3519	9	HSL81689	L81689 Homo sapien
c 135	19	2.3	170073	2	AL845419	AL845419 Danio rer	c 208	18	2.1	3660	9	HSL81688	L81688 Homo sapien
c 136	19	2.3	170967	2	AC005023	AC005023 Homo sapi	c 209	18	2.1	3947	8	NTA319872	X319872 Nicotiana
c 137	19	2.3	171413	9	AC069020	AC069020 Homo sapi	c 210	18	2.1	4673	3	PFASSA	MB4732 Plasmodium
c 138	19	2.3	171915	9	AF246928	AF246928 Homo sapi	c 211	18	2.1	5123	2	AC014388	AC014388 Drosophill

c 212	18	2.1	5768	6	AX347064	AX347064 Sequence	285	18	2.1	100000	9	AP0000067	AP0000067 Homo sapi
c 213	18	2.1	6351	6	AX281476	AX281476 Sequence	286	18	2.1	100528	2	RN129P15	AL671850 Rattus no
c 214	18	2.1	6351	6	AX346647	AX346647 Sequence	287	18	2.1	100635	8	AP003446	Oryza sat
c 215	18	2.1	6351	6	AX348810	AX348810 Sequence	288	18	2.1	101786	8	ATT14K23	AL132909 Arabidops
c 216	18	2.1	6617	6	AX139467	AX139467 Sequence	289	18	2.1	104008	2	AC084847	Homo sapi
c 217	18	2.1	6904	6	AX345114	AX345114 Sequence	290	18	2.1	104533	2	CNS07YPU	AL731760 Oryza sat
c 218	18	2.1	8062	1	AF292097	AF292097 Lepidopir	c 291	18	2.1	106073	9	AL136146	Human DNA
c 219	18	2.1	10352	8	MTCCGNS	Z48930 C.caldarium	c 292	18	2.1	107103	9	AL589702	Human DNA
c 220	18	2.1	11101	8	SCYDR035W	Z74331 S.cerevisia	c 293	18	2.1	107607	9	AP000641	Homo sapi
c 221	18	2.1	12257	1	AB008615	AB008615 Rickettsi	294	18	2.1	108799	9	AC090989	Homo sapi
c 222	18	2.1	12706	8	SCYDR034C	Z74330 S.cerevisia	295	18	2.1	109184	2	AL732394	Homo sapi
c 223	18	2.1	16840	5	XNUL17896	U17896 Xiphophorus	296	18	2.1	109642	2	AL119734	Homo sapi
c 224	18	2.1	20570	6	AX059489	AX059489 Sequence	297	18	2.1	110695	9	AC010306	Homo sapi
c 225	18	2.1	20941	1	AB000976	AB000976 Archaeogl	298	18	2.1	111301	8	AC011717	Arabidops
c 226	18	2.1	22157	6	AX283722	AX283722 Sequence	299	18	2.1	111569	2	AC126697	Rattus no
c 227	18	2.1	24628	8	SC9673	Z68196 S.cerevisia	c 300	18	2.1	111864	9	AP003042	Homo sapi
c 228	18	2.1	24665	9	AL450488	AL450488 Human DNA	c 301	18	2.1	112353	9	AC109928	Homo sapi
c 229	18	2.1	27227	8	SPAC22H10	Z59730 S.pombe chr	c 302	18	2.1	112670	9	AL353894	Human DNA
c 230	18	2.1	29731	3	CEB0391	Z81454 Caenorhabdi	c 303	18	2.1	114119	2	AC112288	Rattus no
c 231	18	2.1	30098	3	CEK0929	Z79602 Caenorhabdi	c 304	18	2.1	114719	2	AC095511	Rattus no
c 232	18	2.1	32179	3	AF047663	AF047663 Caenorhab	c 305	18	2.1	115278	9	AL139819	Human DNA
c 233	18	2.1	36759	3	CET05F1	Z81586 Caenorhabdi	c 306	18	2.1	115884	2	AC105777	Rattus no
c 234	18	2.1	36790	2	AC131448	AC131448 Strongylo	c 307	18	2.1	116126	2	AC024356	Mus muscu
c 235	18	2.1	38545	10	AL669832	AL669832 Mouse DNA	c 308	18	2.1	116679	2	AC123416	Rattus no
c 236	18	2.1	40481	3	CEC28D4	Z82259 Caenorhabdi	c 309	18	2.1	118008	2	AC079410	Homo sapi
c 237	18	2.1	40498	2	AC087659	AC087659 Homo sapi	c 310	18	2.1	118603	9	AC091819	Homo sapi
c 238	18	2.1	40891	9	AL365321	AL365321 Human DNA	c 311	18	2.1	118869	9	AC004903	Homo sapi
c 239	18	2.1	42446	8	ATL73C19	AL050400 Arabidops	c 312	18	2.1	119343	9	AL355595	Human DNA
c 240	18	2.1	43649	2	AC120006	AC120006 Mus muscu	c 313	18	2.1	119875	9	AC108051	Homo sapi
c 241	18	2.1	44879	8	AC026238	AC026238 Arabidops	c 314	18	2.1	119955	9	AL331595	Human DNA
c 242	18	2.1	49376	2	AC023803	AC023803 Mus muscu	c 315	18	2.1	120129	9	AC091560	Human DNA
c 243	18	2.1	51953	2	AC083779	AC083779 Homo sapi	c 316	18	2.1	120405	2	AC046147	Mus muscu
c 244	18	2.1	52433	2	AC101977	AC101977 Arabidops	c 317	18	2.1	120718	2	AC091249	Mus muscu
c 245	18	2.1	54632	8	AB015472	AB015472 Arabidops	c 318	18	2.1	121600	2	CNS07YPU	AL713954 Oryza sat
c 246	18	2.1	54710	2	AC131502	AC131502 Lytechinu	c 319	18	2.1	122752	9	AL356863	Human DNA
c 247	18	2.1	56870	9	AL512822	AL512822 Human DNA	c 320	18	2.1	122974	2	AC025114	Homo sapi
c 248	18	2.1	61967	2	AC110153	AC110153 Mus muscu	c 321	18	2.1	123722	9	AC018515	Homo sapi
c 249	18	2.1	62072	2	AC099850	AC099850 Homo sapi	c 322	18	2.1	124191	9	AC005153	Homo sapi
c 250	18	2.1	63835	2	AC090861	AC090861 Homo sapi	c 323	18	2.1	124324	2	AL158138	Homo sapi
c 251	18	2.1	64165	9	AF288992	AF288992 Homo sapi	c 324	18	2.1	124603	9	AC093119	Homo sapi
c 252	18	2.1	65221	2	AC009693	AC009693 Homo sapi	c 325	18	2.1	127201	2	AC093119	Homo sapi
c 253	18	2.1	65746	2	AC101374	AC101374 Mus muscu	c 326	18	2.1	127701	2	AC121368	Rattus no
c 254	18	2.1	66046	2	AC113935	AC113935 Mus muscu	c 327	18	2.1	127867	9	HS694B14	Human DNA
c 255	18	2.1	66346	2	AC123615	AC123615 Mus muscu	c 328	18	2.1	128001	2	AL732608	Mus muscu
c 256	18	2.1	67544	2	AC100208	AC100208 Mus muscu	c 329	18	2.1	128189	9	AL732509	Human DNA
c 257	18	2.1	67573	2	AC122760	AC122760 Mus muscu	c 330	18	2.1	129849	9	HS710L4	Human DNA
c 258	18	2.1	67851	2	AC117705	AC117705 Mus muscu	c 331	18	2.1	130221	2	AC125495	Oryza sat
c 259	18	2.1	67992	2	AC114573	AC114573 Mus muscu	c 332	18	2.1	130416	9	AC088851	Homo sapi
c 260	18	2.1	68245	2	AC121261	AC121261 Mus muscu	c 333	18	2.1	130935	9	AL596218	Human DNA
c 261	18	2.1	69236	2	AC100226	AC100226 Mus muscu	c 334	18	2.1	131120	2	AC106359	Rattus no
c 262	18	2.1	70216	2	AC034253	AC034253 Homo sapi	c 335	18	2.1	131227	2	AL136098	Homo sapi
c 263	18	2.1	72639	2	AC120190	AC120190 Mus muscu	c 336	18	2.1	131347	2	AC002421	Homo sapi
c 264	18	2.1	74996	9	AP001986	AP001986 Homo sapi	c 337	18	2.1	131395	9	AC084740	Homo sapi
c 265	18	2.1	76187	2	AC015219	AC015219 Drosophil	c 338	18	2.1	131489	9	AL592104	Human DNA
c 266	18	2.1	79949	9	AL162613	AL162613 Human DNA	c 339	18	2.1	131663	2	AC105684	Rattus no
c 267	18	2.1	81513	8	F14D7	AC021198 Sequence	c 340	18	2.1	133630	2	AC117337	Rattus no
c 268	18	2.1	81580	8	AP000736	AP000736 Arabidops	c 341	18	2.1	134558	2	AC117166	Rattus no
c 269	18	2.1	83594	8	AB008268	AB008268 Arabidops	c 342	18	2.1	135458	2	AC117166	Rattus no
c 270	18	2.1	84562	9	AB008725	AB008725 Homo sapi	c 343	18	2.1	135753	2	AC095725	Rattus no
c 271	18	2.1	85195	9	AL135093	AL135093 Human DNA	c 344	18	2.1	136232	2	AC116850	Mus muscu
c 272	18	2.1	85918	2	DNBR33H2	AL121812 Drosophil	c 345	18	2.1	138020	9	AC108043	Homo sapi
c 273	18	2.1	86640	9	HSU50871	HSU50871 Human fapi	c 346	18	2.1	138225	2	AP004114	Oryza sat
c 274	18	2.1	87783	9	AC073055	AC073055 Homo sapi	c 347	18	2.1	138293	2	AC126191	Rattus no
c 275	18	2.1	89453	8	AC068655	AC068655 Genomic S	c 348	18	2.1	138604	9	AC104826	Homo sapi
c 276	18	2.1	90149	8	AC010793	AC010793 Genomic S	c 349	18	2.1	139154	2	AC130893	Homo sapi
c 277	18	2.1	91039	8	AP004491	AP004491 Lotus jap	c 350	18	2.1	139665	2	AP000865	Homo sapi
c 278	18	2.1	93426	9	AL359732	AL359732 Human DNA	c 351	18	2.1	140335	9	AC018645	Homo sapi
c 279	18	2.1	94993	9	AC117534	AC117534 Homo sapi	c 352	18	2.1	141675	2	AP000405	Homo sapi
c 280	18	2.1	95764	2	AC015732	AC015732 Homo sapi	c 353	18	2.1	142672	2	AL355527	Homo sapi
c 281	18	2.1	96547	2	AC128014	AC128014 Rattus no	c 354	18	2.1	142698	9	AC025429	Homo sapi
c 282	18	2.1	96971	2	AL445425	AL445425 Homo sapi	c 355	18	2.1	142949	2	AL162402	Homo sapi
c 283	18	2.1	98017	2	AC027033	AC027033 Arabidops	c 356	18	2.1	143279	9	AC104838	Homo sapi
c 284	18	2.1	98286	9	AC092392	AC092392 Homo sapi	c 357	18	2.1	143299	9	AL353619	Human DNA

C 358	18	2.1 143723	2	AC012248	Homo sapi	C 431	18	2.1 162566	9	AC103724	Homo sapi
C 359	18	2.1 143857	9	AC021583	Homo sapi	C 432	18	2.1 162737	2	AC007733	Homo sapi
C 360	18	2.1 144388	2	AC091433	Homo sapi	C 433	18	2.1 162925	2	AF004461	Oryza sat
C 361	18	2.1 144445	2	AC102291	Mus muscu	C 434	18	2.1 162986	9	CNS05TBQ	Human chr
C 362	18	2.1 145832	2	AC090736	Homo sapi	C 435	18	2.1 163186	9	AB045360	Homo sapi
C 363	18	2.1 145832	2	AC090736	Homo sapi	C 436	18	2.1 163186	9	AB045360	Homo sapi
C 364	18	2.1 145948	10	AL8333785	Mouse DNA	C 437	18	2.1 164275	2	AC021634	Homo sapi
C 365	18	2.1 146957	9	AC084364	Homo sapi	C 438	18	2.1 164295	2	AC025958	Homo sapi
C 366	18	2.1 147722	2	AC016783	Homo sapi	C 439	18	2.1 164354	9	AC022721	Homo sapi
C 367	18	2.1 147859	9	AC092293	Homo sapi	C 440	18	2.1 164931	2	AC021692	Homo sapi
C 368	18	2.1 147940	2	AC026913	Homo sapi	C 441	18	2.1 164973	9	AC087897	Homo sapi
C 369	18	2.1 148018	2	AC068861	Homo sapi	C 442	18	2.1 164984	2	AC117978	Rattus no
C 370	18	2.1 148466	2	AC023167	Mus muscu	C 443	18	2.1 164984	2	AC020986	Homo sapi
C 371	18	2.1 149700	2	AP002738	Homo sapi	C 444	18	2.1 165245	2	AC027599	Homo sapi
C 372	18	2.1 149865	8	OSJN00086		C 445	18	2.1 165443	9	AC0079939	Mus muscu
C 373	18	2.1 150933	2	AC010829	Homo sapi	C 446	18	2.1 165643	9	AC006222	Homo sapi
C 374	18	2.1 150933	2	AC010829	Homo sapi	C 447	18	2.1 165643	9	AC006222	Homo sapi
C 375	18	2.1 151044	2	AC108712	Homo sapi	C 448	18	2.1 166104	2	AC108586	Rattus no
C 376	18	2.1 151044	2	AC108712	Homo sapi	C 449	18	2.1 166104	2	AC108586	Rattus no
C 377	18	2.1 151544	9	AC013509	Homo sapi	C 450	18	2.1 166451	8	AC084405	Oryza sat
C 378	18	2.1 152000	9	AC096644	Homo sapi	C 451	18	2.1 166697	2	AC116205	Rattus no
C 379	18	2.1 152228	9	AL135902	Human DNA	C 452	18	2.1 166771	2	AC098178	Rattus no
C 380	18	2.1 152593	8	AC002332	Arabidops	C 453	18	2.1 167336	2	AC128361	Homo sapi
C 381	18	2.1 153315	2	AC022953	Homo sapi	C 454	18	2.1 167368	9	AP004243	Homo sapi
C 382	18	2.1 153602	2	AC012291	Homo sapi	C 455	18	2.1 167398	9	CNS07EET	Human chr
C 383	18	2.1 153602	2	AC012291	Homo sapi	C 456	18	2.1 167684	2	AC128357	Rattus no
C 384	18	2.1 154053	2	AC110331	Rattus no	C 457	18	2.1 167788	9	AC055714	Homo sapi
C 385	18	2.1 154214	9	AC079753	Homo sapi	C 458	18	2.1 167889	2	AC073354	Homo sapi
C 386	18	2.1 155609	9	HSJ813016	Human DNA	C 459	18	2.1 167904	9	AL157830	Human DNA
C 387	18	2.1 155667	2	AC018619	Homo sapi	C 460	18	2.1 167939	2	AC124896	Rattus no
C 388	18	2.1 155803	9	AC090799	Homo sapi	C 461	18	2.1 167986	2	AC120994	Rattus no
C 389	18	2.1 155827	9	AC046136	Homo sapi	C 462	18	2.1 168032	2	AC022785	Homo sapi
C 390	18	2.1 155937	2	AC016341	Homo sapi	C 463	18	2.1 168366	2	AC114345	Rattus no
C 391	18	2.1 156848	9	AC097710	Homo sapi	C 464	18	2.1 168580	2	AF005297	Oryza sat
C 392	18	2.1 156992	9	AC093581	Homo sapi	C 465	18	2.1 168621	9	AL157830	Human DNA
C 393	18	2.1 157356	2	AC021376	Homo sapi	C 466	18	2.1 168862	2	AC130401	Homo sapi
C 394	18	2.1 157377	2	AC022827	Homo sapi	C 467	18	2.1 168866	2	AC115075	Mus muscu
C 395	18	2.1 157546	2	AC019181	Mus muscu	C 468	18	2.1 168903	2	AL844187	Danio rer
C 396	18	2.1 157633	2	AC114863	Rattus no	C 469	18	2.1 169258	9	AC022013	Homo sapi
C 397	18	2.1 158094	2	AC108623	Rattus no	C 470	18	2.1 169293	2	AC111249	Rattus no
C 398	18	2.1 158216	2	AC024463	Homo sapi	C 471	18	2.1 169408	2	AC079258	Homo sapi
C 399	18	2.1 158456	2	CNS01DXB	Human chr	C 472	18	2.1 169547	2	AC095580	Rattus no
C 400	18	2.1 158508	2	AC063934	Homo sapi	C 473	18	2.1 169751	2	AC130785	Papio cyn
C 401	18	2.1 158566	2	AC093694	Homo sapi	C 474	18	2.1 169828	9	AC097461	Homo sapi
C 402	18	2.1 158981	2	AC016242	Homo sapi	C 475	18	2.1 169988	9	AC064846	Homo sapi
C 403	18	2.1 159007	9	AC020658	Homo sapi	C 476	18	2.1 170367	2	AC025804	Homo sapi
C 404	18	2.1 159285	9	AL354816	Human DNA	C 477	18	2.1 170496	2	AL19342	Rattus no
C 405	18	2.1 159508	2	AC018573	Homo sapi	C 478	18	2.1 170623	9	AL391375	Human DNA
C 406	18	2.1 159561	2	AC117307	Rattus no	C 479	18	2.1 170634	2	AC034163	Homo sapi
C 407	18	2.1 159697	9	AC012354	Oryza sat	C 480	18	2.1 170644	2	AC026269	Homo sapi
C 408	18	2.1 159739	2	AC117768	Mus muscu	C 481	18	2.1 170834	2	AC021792	Homo sapi
C 409	18	2.1 159859	2	AC074010	Homo sapi	C 482	18	2.1 170970	9	AL512510	Human DNA
C 410	18	2.1 159908	2	AC117911	Rattus no	C 483	18	2.1 171031	2	AC104020	Homo sapi
C 411	18	2.1 160023	2	AC121407	Rattus no	C 484	18	2.1 171188	2	AC005025	Homo sapi
C 412	18	2.1 160042	2	AC027757	Homo sapi	C 485	18	2.1 171239	2	AC092873	Pan trogl
C 413	18	2.1 160392	9	AC104236	Homo sapi	C 486	18	2.1 171391	2	AC098646	Pan trogl
C 414	18	2.1 160439	9	AC005345	Homo sapi	C 487	18	2.1 172327	9	AC011716	Homo sapi
C 415	18	2.1 160741	9	AC097709	Homo sapi	C 488	18	2.1 172482	9	AC022309	Homo sapi
C 416	18	2.1 160789	9	AL589655	Human DNA	C 489	18	2.1 172573	9	AC007276	Homo sapi
C 417	18	2.1 160918	2	AC069196	Homo sapi	C 490	18	2.1 172983	9	AC002454	Human BAC
C 418	18	2.1 160986	3	AC108481	Drosophil	C 491	18	2.1 172985	2	AC087695	Homo sapi
C 419	18	2.1 161008	4	AP003255	Oryza sat	C 492	18	2.1 173487	2	AC026925	Homo sapi
C 420	18	2.1 161008	4	AP003255	Oryza sat	C 493	18	2.1 173728	9	AC018507	Homo sapi
C 421	18	2.1 161008	4	AP003255	Oryza sat	C 494	18	2.1 173874	2	AC068011	Homo sapi
C 422	18	2.1 161103	9	AL392163	Human DNA	C 495	18	2.1 174030	2	AC092555	Papio cyn
C 423	18	2.1 161166	2	AC087673	Homo sapi	C 496	18	2.1 174535	9	AC090465	Homo sapi
C 424	18	2.1 161557	2	AC009826	Homo sapi	C 497	18	2.1 174741	2	AC024280	Homo sapi
C 425	18	2.1 161570	2	AC111807	Rattus no	C 498	18	2.1 174803	9	AL591501	Human DNA
C 426	18	2.1 161577	2	AC118924	Mus muscu	C 499	18	2.1 175097	2	AC112639	Rattus no
C 427	18	2.1 162151	9	AC012358	Homo sapi	C 500	18	2.1 175330	9	AC018457	Homo sapi
C 428	18	2.1 162312	2	AC127074	Rattus no	C 501	18	2.1 175493	3	AC009379	Drosophil
C 429	18	2.1 162328	2	AC129300	Mus muscu	C 502	18	2.1 175965	2	AC026176	Homo sapi
C 430	18	2.1 162549	9	AC090802	Homo sapi	C 503	18	2.1 176366	2	AC007941	Homo sapi

504	18	2.1	176619	2	AC026497	Homo sapi	AC026497	18	2.1	190537	2	AC010757	Homo sapi
c 505	18	2.1	176626	9	AC084754	Homo sapi	AC084754	18	2.1	192169	9	AC099560	Homo sapi
c 506	18	2.1	176888	2	AC119081	Rattus no	AC119081	18	2.1	192253	2	AC020615	Homo sapi
c 507	18	2.1	177073	2	AP000934	Homo sapi	AP000934	18	2.1	192330	2	AC123157	Canis fam
c 508	18	2.1	177074	9	AP000487	Homo sapi	AP000487	18	2.1	192339	2	AC128366	Rattus no
c 509	18	2.1	177852	2	AC034283	Homo sapi	AC034283	18	2.1	192359	2	AC073855	Homo sapi
c 510	18	2.1	178175	2	AC124067	Homo sapi	AC124067	18	2.1	193024	9	AL139275	Human DNA
c 511	18	2.1	178194	2	AC114216	Rattus no	AC114216	18	2.1	193126	9	AL139275	Human DNA
c 512	18	2.1	178825	2	AP001203	Homo sapi	AP001203	18	2.1	193300	2	AL139275	Human DNA
c 513	18	2.1	178870	2	AC021555	Homo sapi	AC021555	18	2.1	193364	2	AC108424	Mus muscu
c 514	18	2.1	179040	2	AC022732	Homo sapi	AC022732	18	2.1	193563	9	AC005538	Homo sapi
c 515	18	2.1	179536	2	AC010814	Homo sapi	AC010814	18	2.1	193650	9	AC093870	Homo sapi
c 516	18	2.1	179617	2	AC073022	Homo sapi	AC073022	18	2.1	193928	2	AL772147	Mus muscu
c 517	18	2.1	179695	2	H8512L9	Homo sapi	AL445624	18	2.1	194304	9	AC016708	Homo sapi
c 518	18	2.1	180122	2	AC101949	Mus muscu	AC101949	18	2.1	194467	2	AC126721	Rattus no
c 519	18	2.1	180209	2	AL512587	Mus muscu	AL512587	18	2.1	194590	2	AL645962	Mus muscu
c 520	18	2.1	180387	3	AC009378	Drosophil	AC009378	18	2.1	194757	2	AC068882	Homo sapi
c 521	18	2.1	180424	9	AC106860	Homo sapi	AC106860	18	2.1	194974	2	AC107787	Mus muscu
c 522	18	2.1	180487	2	AC113228	Homo sapi	AC113228	18	2.1	195224	2	AC120767	Rattus no
c 523	18	2.1	180625	9	AL391478	Human DNA	AL391478	18	2.1	195316	2	AC044918	Homo sapi
c 524	18	2.1	180650	9	AC109474	Homo sapi	AC109474	18	2.1	195616	2	AC090344	Homo sapi
c 525	18	2.1	180685	9	AC006432	Homo sapi	AC006432	18	2.1	195725	2	AC024115	Mus muscu
c 526	18	2.1	180716	2	AC126876	Rattus no	AC126876	18	2.1	196112	2	AC112067	Rattus no
c 527	18	2.1	180892	2	AC097732	Rattus no	AC097732	18	2.1	196344	2	AC098594	Homo sapi
c 528	18	2.1	180892	2	AC097732	Rattus no	AC097732	18	2.1	196552	2	AC126537	Rattus no
c 529	18	2.1	181222	2	AC122500	Mus muscu	AC122500	18	2.1	196552	2	AC126537	Rattus no
c 530	18	2.1	181356	2	AC110210	Mus muscu	AC110210	18	2.1	196606	9	AP000779	Homo sapi
c 531	18	2.1	181782	2	AC022312	Homo sapi	AC022312	18	2.1	196671	9	AC073479	Homo sapi
c 532	18	2.1	181782	2	AC015566	Homo sapi	AC015566	18	2.1	196756	2	AC121858	Mus muscu
c 533	18	2.1	181878	2	AC126027	Mus muscu	AC126027	18	2.1	196756	2	AC026775	Homo sapi
c 534	18	2.1	181930	2	AC125614	Rattus no	AC125614	18	2.1	196836	9	AC090345	Homo sapi
c 535	18	2.1	181949	9	AC006196	Homo sapi	AC006196	18	2.1	197156	2	AC090345	Homo sapi
c 536	18	2.1	181985	2	AC096382	Rattus no	AC096382	18	2.1	197226	2	AC116593	Mus muscu
c 537	18	2.1	182778	2	AC022312	Homo sapi	AC022312	18	2.1	197236	9	AC018710	Homo sapi
c 538	18	2.1	183186	9	AC093817	Homo sapi	AC093817	18	2.1	197346	2	AC095701	Rattus no
c 539	18	2.1	183334	2	AC106174	Rattus no	AC106174	18	2.1	197455	9	AC005701	Homo sapi
c 540	18	2.1	183337	2	AL139002	Human DNA	AL139002	18	2.1	197781	2	AC094493	Rattus no
c 541	18	2.1	183370	2	AL844531	Mus muscu	AL844531	18	2.1	198006	2	AL731852	Mus muscu
c 542	18	2.1	183975	10	AL590988	Mouse DNA	AL590988	18	2.1	198412	2	AC098236	Rattus no
c 543	18	2.1	184104	9	AC020709	Homo sapi	AC020709	18	2.1	198422	2	AC113939	Mus muscu
c 544	18	2.1	184455	2	AC095277	Rattus no	AC095277	18	2.1	198777	8	ATCHRIV63	Arabidops
c 545	18	2.1	184620	2	AC096320	Rattus no	AC096320	18	2.1	199200	9	AC010351	Homo sapi
c 546	18	2.1	185125	2	AC101663	Mus muscu	AC101663	18	2.1	199612	9	AC021582	Homo sapi
c 547	18	2.1	185650	2	AC122338	Mus muscu	AC122338	18	2.1	199814	3	CEY47D3A	Caenorhab
c 548	18	2.1	185829	2	AC094911	Rattus no	AC094911	18	2.1	200105	9	AC026975	Homo sapi
c 549	18	2.1	185923	2	AC129069	Papio cyn	AC129069	18	2.1	200539	2	AC121805	Mus muscu
c 550	18	2.1	185995	2	AC087306	Homo sapi	AC087306	18	2.1	200567	2	AP001587	Homo sapi
c 551	18	2.1	186074	9	AP005061	Homo sapi	AP005061	18	2.1	201312	2	AC067900	Homo sapi
c 552	18	2.1	186077	2	AC111573	Rattus no	AC111573	18	2.1	201886	2	AC129181	Mus muscu
c 553	18	2.1	186091	9	AC011242	Homo sapi	AC011242	18	2.1	201989	2	AC073169	Homo sapi
c 554	18	2.1	186105	2	AC010784	Homo sapi	AC010784	18	2.1	202317	9	AC020603	Homo sapi
c 555	18	2.1	186333	2	AC027355	Mus muscu	AC027355	18	2.1	202873	2	AC104895	Mus muscu
c 556	18	2.1	187038	2	AC021179	Homo sapi	AC021179	18	2.1	203405	2	AC024964	Homo sapi
c 557	18	2.1	187072	9	AC099541	Homo sapi	AC099541	18	2.1	203540	2	AP002396	Homo sapi
c 558	18	2.1	187175	2	AC108832	Mus muscu	AC108832	18	2.1	203741	2	AC117585	Mus muscu
c 559	18	2.1	187193	9	AC007422	Homo sapi	AC007422	18	2.1	204117	2	AC023045	Homo sapi
c 560	18	2.1	187275	9	AC107447	Rattus no	AC107447	18	2.1	204132	2	AC121396	Rattus no
c 561	18	2.1	187318	2	AC112939	Mus muscu	AC112939	18	2.1	204322	2	AC027648	Mus muscu
c 562	18	2.1	187343	2	AC016688	Homo sapi	AC016688	18	2.1	204784	2	AC091547	Rattus no
c 563	18	2.1	187439	2	AC107710	Mus muscu	AC107710	18	2.1	205557	2	AC102170	Mus muscu
c 564	18	2.1	187488	2	AC010454	Homo sapi	AC010454	18	2.1	206214	2	AC079859	Homo sapi
c 565	18	2.1	187837	2	AC130991	Rattus no	AC130991	18	2.1	206240	10	AL772190	Mouse DNA
c 566	18	2.1	187889	9	AC093008	Homo sapi	AC093008	18	2.1	207737	10	AL589679	Mouse DNA
c 567	18	2.1	188047	2	AC022995	Homo sapi	AC022995	18	2.1	209877	9	AC012509	Homo sapi
c 568	18	2.1	188356	2	AC009550	Homo sapi	AC009550	18	2.1	210595	2	AC097034	Rattus no
c 569	18	2.1	188719	9	AC048330	Homo sapi	AC048330	18	2.1	210790	2	AC126038	Mus muscu
c 570	18	2.1	189224	2	AC122958	Rattus no	AC122958	18	2.1	211998	2	AL732530	Mus muscu
c 571	18	2.1	189445	2	AC018739	Homo sapi	AC018739	18	2.1	212693	2	AL773516	Mus muscu
c 572	18	2.1	190054	2	AC114730	Homo sapi	AC114730	18	2.1	212810	2	AL772201	Mus muscu
c 573	18	2.1	190235	2	AC118202	Mus muscu	AC118202	18	2.1	213743	2	AL359918	Homo sapi
c 574	18	2.1	190418	2	AP000777	Homo sapi	AP000777	18	2.1	214727	2	AC094443	Rattus no
c 575	18	2.1	190462	9	AP003057	Homo sapi	AP003057	18	2.1	215738	2	AC113274	Mus muscu
c 576	18	2.1	190493	2	AC118397	Rattus no	AC118397	18	2.1	215874	2	AC114557	Mus muscu
								18	2.1	216100	2	AC129180	Mus muscu
								18	2.1	216250	9	AC099731	Homo sapi

C 650	18	2.1	218730	2	AC090571	AC090571 Homo sapi	C 723	17	2.0	543	6	AX432125	Sequence
C 651	18	2.1	219013	9	AC008731	AC008731 Homo sapi	C 724	17	2.0	545	6	AX387845	Sequence
C 652	18	2.1	220183	9	AC011246	AC011246 Homo sapi	C 725	17	2.0	555	6	AX099149	Sequence
C 653	18	2.1	221004	9	CNS01DRW	AL121576 Human chr	C 726	17	2.0	557	4	DRCPRP1A	LJ5449 Dromaciops
C 654	18	2.1	221375	2	AC124710	AC124710 Mus muscu	C 727	17	2.0	567	8	MCO286644	AF286644 Macrozami
C 655	18	2.1	223088	2	AC099383	AC099383 Rattus no	C 728	17	2.0	623	1	AF274738	AF274738 Enterococ
C 656	18	2.1	225090	2	AC094399	AC094399 Rattus no	C 729	17	2.0	623	6	AX109993	Sequence
C 657	18	2.1	225208	2	AC015561	AC015561 Homo sapi	C 730	17	2.0	627	1	NME270898	AF270898 Neisseria
C 658	18	2.1	231322	2	AC022775	AC022775 Mus muscu	C 731	17	2.0	627	6	AR099157	Sequence
C 659	18	2.1	231430	2	AC069063	AC069063 Homo sapi	C 732	17	2.0	630	1	NLA270904	AF270904 Neisseria
C 660	18	2.1	234189	2	AC125040	AC125040 Mus muscu	C 733	17	2.0	630	1	NME270887	AF270887 Neisseria
C 661	18	2.1	238331	2	AC099071	AC099071 Rattus no	C 734	17	2.0	630	1	NME270888	AF270888 Neisseria
C 662	18	2.1	240342	2	AC092862	AC092862 Homo sapi	C 735	17	2.0	630	1	NME270914	AF270914 Neisseria
C 663	18	2.1	241543	2	AC091278	AC091278 Mus muscu	C 736	17	2.0	632	11	G62295	G62295 B257N2/SP6
C 664	18	2.1	242730	2	AC099408	AC099408 Pan trogl	C 737	17	2.0	642	1	AY034824	AY034824 Haemophil
C 665	18	2.1	243382	2	AC094242	AC094242 Rattus no	C 738	17	2.0	645	9	HST28C	X05930 Human DNA r
C 666	18	2.1	246823	2	AC079483	AC079483 Mus muscu	C 739	17	2.0	663	6	AX139250	AX139250 Sequence
C 667	18	2.1	253389	2	AC008605	AC008605 Homo sapi	C 740	17	2.0	700	6	AX182694	AX182694 Sequence
C 668	18	2.1	253662	2	AC122213	AC122213 Mus muscu	C 741	17	2.0	700	6	HSA326761	AF326761 Homo sapi
C 669	18	2.1	258657	2	AC094700	AC094700 Rattus no	C 742	17	2.0	754	11	AF205819	AF205819 Homo sapi
C 670	18	2.1	262860	2	AC067970	AC067970 Mus muscu	C 743	17	2.0	757	8	AF064534	AF064534 Candida a
C 671	18	2.1	266050	1	AP000988	AP000988 Sulfolobu	C 744	17	2.0	802	6	AX478112	AX478112 Sequence
C 672	18	2.1	272850	1	AP004828	AP004828 Staphyloc	C 745	17	2.0	867	1	AF368004	AF368004 Cowdria r
C 673	18	2.1	278284	2	AC095376	AC095376 Rattus no	C 746	17	2.0	824	11	CNS06103	AL399673 T7 end of
C 674	18	2.1	279530	3	AE003518	AE003518 Drosophi	C 747	17	2.0	831	1	HIHIFA	X16991 Haemophilus
C 675	18	2.1	280810	2	AC046188	AC046188 Homo sapi	C 748	17	2.0	838	9	HST28A	X05928 Human germl
C 676	18	2.1	282368	2	AC126684	AC126684 Mus muscu	C 749	17	2.0	872	1	HIU19761	X09761 Haemophilus
C 677	18	2.1	291150	1	AP003135	AP003135 Staphyloc	C 750	17	2.0	885	1	HIPIII	X52419 H.influenza
C 678	18	2.1	294218	3	AE003430	AE003430 Drosophi	C 751	17	2.0	888	8	AF320263	AF320263 Glycosmis
C 679	18	2.1	295409	3	AC127976	AC127976 Rattus no	C 752	17	2.0	994	11	CNS06KP2	AL403164 T3 end of
C 680	18	2.1	299864	2	AC006702	AC006702 Caenorhab	C 753	17	2.0	1044	8	CABCH2	Y09722 C.annuum mR
C 681	18	2.1	300000	9	AP002533	AP002533 Homo sapi	C 754	17	2.0	1139	6	AG2799	AG2799 Sequence 40
C 682	18	2.1	310050	1	RM6603642	AL603642 Rhizobium	C 755	17	2.0	1139	9	HSBRCA224	X95174 H.sapiens b
C 683	18	2.1	312430	1	RPXX02	AL235271 Rickettsi	C 756	17	2.0	1158	3	AY094848	AY094848 Drosophi
C 684	18	2.1	321304	3	CEY105C5B	AL110479 Caenorhab	C 757	17	2.0	1164	6	AG2798	AG2798 Sequence 39
C 685	18	2.1	323394	1	AP003194	AP003194 Clostridi	C 758	17	2.0	1164	9	HSBRCA223	X95173 H.sapiens b
C 686	18	2.1	335259	2	AC015834	AC015834 Homo sapi	C 759	17	2.0	1199	6	AG2797	AG2797 Sequence 38
C 687	18	2.1	340000	9	AP001680	AP001680 Homo sapi	C 760	17	2.0	1199	9	HSBRCA222	X95172 H.sapiens b
C 688	18	2.1	342600	1	AP003363	AP003363 Staphyloc	C 761	17	2.0	1210	5	AF084948	AF084948 Dantio rer
C 689	17	2.0	24	6	AG2910	AG2910 Sequence 15	C 762	17	2.0	1252	9	HSSAD5H1	AF009736 Homo sapi
C 690	17	2.0	65	6	AX485579	AX485579 Sequence	C 763	17	2.0	1384	3	DMU13068	U13068 Drosophila
C 691	17	2.0	75	6	AR099154	AR099154 Sequence	C 764	17	2.0	1467	1	CRMAPI	X74250 C.ruminanti
C 692	17	2.0	90	6	AR099209	AR099209 Sequence	C 765	17	2.0	1500	1	MBU38919	MBU38919 Methanosarc
C 693	17	2.0	90	6	AR154303	AR154303 Sequence	C 766	17	2.0	1503	8	AF370603	AF370603 Arabidops
C 694	17	2.0	177	6	AR184447	AR184447 Sequence	C 767	17	2.0	1518	9	BC026098	BC026098 Homo sapi
C 695	17	2.0	216	9	AY008851	AY008851 Homo sapi	C 768	17	2.0	1524	10	AF241257	AF241257 Mus muscu
C 696	17	2.0	256	11	DM118C11S	Z32429 D. melanoga	C 769	17	2.0	1574	9	HSB03827	AL832519 Homo sapi
C 697	17	2.0	276	6	AR099189	AR099189 Sequence	C 770	17	2.0	1574	9	HSB03827	AL832519 Homo sapi
C 698	17	2.0	310	6	AX340463	AX340463 Sequence	C 771	17	2.0	1706	6	AX364087	AX364087 Sequence
C 699	17	2.0	331	11	G35765	G35765 SRG hl4a975	C 772	17	2.0	1706	6	AX366173	AX366173 Sequence
C 700	17	2.0	357	6	AR099188	AR099188 Sequence	C 773	17	2.0	1721	8	AF419579	AF419579 Arabidops
C 701	17	2.0	402	6	AR099183	AR099183 Sequence	C 774	17	2.0	1742	8	AB072477	AB072477 Lycopersi
C 702	17	2.0	424	11	AU028501	AU028501 Rattus no	C 775	17	2.0	1748	8	SCYKLI07W	Z28107 S.cerevisia
C 703	17	2.0	426	6	AR099202	AR099202 Sequence	C 776	17	2.0	1772	9	AX093765	AX093765 Homo sapi
C 704	17	2.0	426	6	AR099217	AR099217 Sequence	C 777	17	2.0	1830	8	AY042900	AY042900 Arabidops
C 705	17	2.0	426	6	AR099223	AR099223 Sequence	C 778	17	2.0	1833	1	AF194870	AF194870 Clostridi
C 706	17	2.0	426	6	AR099226	AR099226 Sequence	C 779	17	2.0	1933	8	CAU07366	U07366 Candida alb
C 707	17	2.0	426	6	AR099229	AR099229 Sequence	C 780	17	2.0	2044	6	AR037680	AR037680 Sequence
C 708	17	2.0	426	6	AR099232	AR099232 Sequence	C 781	17	2.0	2044	6	AR120647	AR120647 Sequence
C 709	17	2.0	426	6	AR099235	AR099235 Sequence	C 782	17	2.0	2044	6	AR120648	AR120648 Sequence
C 710	17	2.0	426	6	AR099238	AR099238 Sequence	C 783	17	2.0	2088	9	HSM805186	AL833887 Homo sapi
C 711	17	2.0	426	6	AR099243	AR099243 Sequence	C 784	17	2.0	2106	6	AX489100	AX489100 Sequence
C 712	17	2.0	426	6	AR154296	AR154296 Sequence	C 785	17	2.0	2280	8	AF078874	AF078874 Avena sat
C 713	17	2.0	426	6	AR154311	AR154311 Sequence	C 786	17	2.0	2281	8	AF078873	AF078873 Avena sat
C 714	17	2.0	444	6	AR099182	AR099182 Sequence	C 787	17	2.0	2457	10	BC016599	BC016599 Mus muscu
C 715	17	2.0	516	6	AR099220	AR099220 Sequence	C 788	17	2.0	2520	1	AF085340	AF085340 Neisseria
C 716	17	2.0	518	11	HSPE46F04	AL010044 H.sapiens	C 789	17	2.0	2520	6	SCHYPI	X56235 Yeast (S.ce
C 717	17	2.0	525	6	AR099160	AR099160 Sequence	C 790	17	2.0	2538	8	AX370172	AX370172 Sequence
C 718	17	2.0	525	6	AR099181	AR099181 Sequence	C 791	17	2.0	2623	6	AX370172	AX370172 Sequence
C 719	17	2.0	525	6	AR099190	AR099190 Sequence	C 792	17	2.0	2623	10	MMGXIN1	U33275 Mus musculu
C 720	17	2.0	525	6	AR099191	AR099191 Sequence	C 793	17	2.0	2640	10	AF167403	AF167403 Mus muscu
C 721	17	2.0	531	9	AF417137	AF417137 Homo sapi	C 794	17	2.0	2664	6	E06740	E06740 Rhizopus ni
C 722	17	2.0	532	9	HUMCMP4	M55678 Human carti	C 795	17	2.0	2670	3	AF070689	AF070689 Drosophi
										2704	3	AF070692	Drosophi



C 796	17	2.0	2767	8	AF526260	Glycine m	AF526260	869	17	2.0	13882	1	AE007576	Clostridi
C 797	17	2.0	2806	10	AF299345	Mus muscu	AF299345	C 870	17	2.0	14117	6	AX331935	Sequence
C 798	17	2.0	2833	10	BC005414	Mus muscu	BC005414	C 871	17	2.0	14117	9	HS8HA2GEN	X90761 Homo sapien
C 799	17	2.0	2841	10	AB028895	Mus muscu	AB028895	C 872	17	2.0	14588	2	AC017550	Drosophil
C 800	17	2.0	2859	3	AF263371	Drosophil	AF263371	C 873	17	2.0	14910	4	SSRYRCRC1	Z54280 S. afroga ge
C 801	17	2.0	2870	3	DMU55767	Drosophila	DMU55767	C 874	17	2.0	15122	3	CEH40L08	Z53310 Caenorhabdi
C 802	17	2.0	2979	3	AF037197	Heliothis	AF037197	C 875	17	2.0	15157	9	AF0000259	AE000259 Homo sapi
C 803	17	2.0	2989	3	HELEST7	Heliothis v	HELEST7	C 876	17	2.0	15451	1	AE0006677	AE000677 Aquifex a
C 804	17	2.0	2989	6	I50971	Sequence 1	I50971	C 877	17	2.0	16120	2	AC015279	AC015279 Drosophil
C 805	17	2.0	2989	6	I67826	Sequence 1	I67826	C 878	17	2.0	17048	5	D21070	D21070 Rana catesb
C 806	17	2.0	2989	6	I68136	Sequence 1	I68136	C 879	17	2.0	17727	3	U61945	U61945 Caenorhabdi
C 807	17	2.0	2990	8	RICPHD2	Oriza sativ	RICPHD2	C 880	17	2.0	17959	6	AX281300	AX281300 Sequence
C 808	17	2.0	2999	3	AF037196	Heliothis	AF037196	C 881	17	2.0	17959	6	AX345477	AX345477 Sequence
C 809	17	2.0	3021	3	AF189719	Giardia 1	AF189719	C 882	17	2.0	18855	6	AX345512	AX345512 Sequence
C 810	17	2.0	3040	6	AR005011	Sequence	AR005011	C 883	17	2.0	19345	9	AL772178	AL772178 Human DNA
C 811	17	2.0	3040	6	AR037062	Sequence	AR037062	C 884	17	2.0	20657	2	AC012869	AC012869 Drosophil
C 812	17	2.0	3040	6	AR082616	Sequence	AR082616	C 885	17	2.0	20820	2	AC014862	AC014862 Drosophil
C 813	17	2.0	3047	6	I50972	Sequence 2	I50972	C 886	17	2.0	20988	9	AL163532	AL163532 Human DNA
C 814	17	2.0	3047	6	I67827	Sequence 2	I67827	C 887	17	2.0	21898	9	AL356576	AL356576 Human DNA
C 815	17	2.0	3047	6	I68137	Sequence 2	I68137	C 888	17	2.0	23889	3	AE002763	AE002763 Drosophil
C 816	17	2.0	3056	8	AF033210	Pneumocys	AF033210	C 889	17	2.0	24346	2	AC116624	AC116624 Homo sapi
C 817	17	2.0	3171	14	MCCP1A	M65139 Cowpea chlo	M65139	C 890	17	2.0	24617	8	AF359360	AF359360 Fusarium
C 818	17	2.0	3246	1	HPKATA	Z70679 H pylori ka	Z70679	C 891	17	2.0	25357	2	AC100975	AC100975 Mus muscu
C 819	17	2.0	3284	3	AF067148	Giardia 1	AF067148	C 892	17	2.0	25389	8	SPBB10D8	SPBB10D8 S. pombe C
C 820	17	2.0	3779	1	CRU74759	Chlamydia t	CRU74759	C 893	17	2.0	25593	3	CEH21P03	CEH21P03 Caenorhabdi
C 821	17	2.0	4007	1	AF283839	Burkholder	AF283839	C 894	17	2.0	26000	9	AF000610	AF000610 Homo sapi
C 822	17	2.0	4397	1	SAPS194	X06627 Staphylococ	X06627	C 895	17	2.0	26402	2	AC015428	AC015428 Drosophil
C 823	17	2.0	4421	9	HUM16DC88Z	L46899 Homo sapien	L46899	C 896	17	2.0	27046	3	AC115581	AC115581 Dictyoste
C 824	17	2.0	5171	8	AF364179	Fusarium	AF364179	C 897	17	2.0	27299	9	AL139810	AL139810 Human DNA
C 825	17	2.0	5190	2	AC014866	Drosophil	AC014866	C 898	17	2.0	27748	2	AC068804	AC068804 Mus muscu
C 826	17	2.0	5504	6	AX251955	Sequence	AX251955	C 899	17	2.0	27878	9	HSJ3E10	HSJ3E10 Human DNA s
C 827	17	2.0	5504	6	AX344347	Sequence	AX344347	C 900	17	2.0	27949	9	AF001135	AF001135 Homo sapi
C 828	17	2.0	5504	6	AX349004	Sequence	AX349004	C 901	17	2.0	28444	2	AC016609	AC016609 Homo sapi
C 829	17	2.0	5534	3	DROECR	M74078 Drosophila	M74078	C 902	17	2.0	28701	9	AC090428	AC090428 Homo sapi
C 830	17	2.0	5871	8	AB001920	Oryza sat	AB001920	C 903	17	2.0	29000	8	SPBC4B4	SPBC4B4 S. pombe C
C 831	17	2.0	5941	14	AF036333	Diatraea	AF036333	C 904	17	2.0	29086	3	CEC05C12	CEC05C12 Caenorhabdi
C 832	17	2.0	6396	6	AX347149	Sequence	AX347149	C 905	17	2.0	29591	9	AB014082	AB014082 Homo sapi
C 833	17	2.0	6814	1	AF013713	Methanosa	AF013713	C 906	17	2.0	29871	1	LLLPK214	LLLPK214 Lactococcus
C 834	17	2.0	6822	8	ZMCIN4E	Y00086 Zea mays DN	Y00086	C 907	17	2.0	30190	9	AC108010	AC108010 Homo sapi
C 835	17	2.0	7744	1	HIFIMGC	Z33502 H. influenza	Z33502	C 908	17	2.0	30200	8	SPAC20G8	SPAC20G8 S. pombe chr
C 836	17	2.0	8067	6	AX344817	Sequence	AX344817	C 909	17	2.0	30336	3	CEBE10	CEBE10 Dictyoste
C 837	17	2.0	8320	2	AC017577	Drosophil	AC017577	C 910	17	2.0	30523	2	AC115592	AC115592 Dictyoste
C 838	17	2.0	8675	4	SSC251837	Sus scrofi	SSC251837	C 911	17	2.0	31028	14	AF391541	AF391541 Bovine co
C 839	17	2.0	8743	2	AC019981	Drosophil	AC019981	C 912	17	2.0	31028	14	AF391542	AF391542 Bovine co
C 840	17	2.0	9500	1	PNFIMOPR	Z32686 Proteus mir	Z32686	C 913	17	2.0	31100	14	AF220295	AF220295 Bovine co
C 841	17	2.0	9590	1	AE013145	Thermoana	AE013145	C 914	17	2.0	31256	9	AC015140	AC015140 Human DNA
C 842	17	2.0	9828	1	AE014084	Buchnera	AE014084	C 915	17	2.0	31505	2	AC015140	AC015140 Drosophil
C 843	17	2.0	10029	1	AE011017	Methanosa	AE011017	C 916	17	2.0	31607	2	AC013178	AC013178 Caenorhabdi
C 844	17	2.0	10029	1	AE013932	Yersinia	AE013932	C 917	17	2.0	31768	3	CET02D1	CET02D1 Caenorhabdi
C 845	17	2.0	10153	1	AE007546	Yersinia	AE007546	C 918	17	2.0	32129	3	CET23F11	CET23F11 Caenorhabdi
C 846	17	2.0	10353	1	AE008328	Agrobacte	AE008328	C 919	17	2.0	32911	3	AF000191	AF000191 Caenorhab
C 847	17	2.0	10454	1	AE010652	Agrobacte	AE010652	C 920	17	2.0	33565	2	AC091830	AC091830 Homo sapi
C 848	17	2.0	10492	1	AE009284	Agrobacte	AE009284	C 921	17	2.0	33780	3	CET05C12	CE05C12 Caenorhabdi
C 849	17	2.0	10511	1	AE008373	Agrobacte	AE008373	C 922	17	2.0	33780	9	AC113553	AC113553 Homo sapi
C 850	17	2.0	10747	3	AE006361	Lactococc	AE006361	C 923	17	2.0	34332	3	CEC33D9	CEC33D9 Caenorhabdi
C 851	17	2.0	10764	3	AE002812	Drosophil	AE002812	C 924	17	2.0	34967	3	CEFI4F3	CEFI4F3 Caenorhabdi
C 852	17	2.0	10807	1	CPU88070	Chlamydomphi	CPU88070	C 925	17	2.0	35757	8	SCAPLAP	SCAPLAP X71133 S. cerevisia
C 853	17	2.0	10891	6	AX345367	Sequence	AX345367	C 926	17	2.0	35800	9	AC005524	AC005524 Homo sapi
C 854	17	2.0	10920	1	AE006362	Lactococc	AE006362	C 927	17	2.0	36306	3	CEM142	CEM142 Caenorhabdi
C 855	17	2.0	10952	1	U02835	Haemophilus	U02835	C 928	17	2.0	36787	2	AC023619	AC023619 Mus muscu
C 856	17	2.0	11034	1	AE004247	Vibrio ch	AE004247	C 929	17	2.0	37115	9	D87016	D87016 Homo sapien
C 857	17	2.0	11130	1	AF312861	Fusobacte	AF312861	C 930	17	2.0	37389	9	AC119040	AC119040 Homo sapi
C 858	17	2.0	11145	1	AE010606	Fusobacte	AE010606	C 931	17	2.0	37858	9	AL731769	AL731769 Human DNA
C 859	17	2.0	11180	1	AE013811	Yersinia	AE013811	C 932	17	2.0	38216	9	HS37B18	HS37B18 Human DNA
C 860	17	2.0	11331	1	AE009240	Agrobacte	AE009240	C 933	17	2.0	38357	9	HS462D8	HS462D8 Human DNA
C 861	17	2.0	11465	14	AF075257	Venezuela	AF075257	C 934	17	2.0	38616	9	AC004208	AC004208 Homo sapi
C 862	17	2.0	11910	1	AE007537	Clostridi	AE007537	C 935	17	2.0	38756	9	D87018	D87018 Homo sapien
C 863	17	2.0	12073	1	AE013919	Yersinia	AE013919	C 936	17	2.0	38784	2	AC014012	AC014012 Drosophil
C 864	17	2.0	12103	10	AF299344	Mus muscu	AF299344	C 937	17	2.0	38953	9	AL360075	AL360075 Human DNA
C 865	17	2.0	13158	6	AF079314	Sequence	AF079314	C 938	17	2.0	38976	3	AF063007	AF063007 Caenorhab
C 866	17	2.0	13285	1	AE009241	Agrobacte	AE009241	C 939	17	2.0	39400	2	AC131245	AC131245 Clona sav
C 867	17	2.0	13631	1	AE007555	Clostridi	AE007555	C 940	17	2.0	39649	3	U41541	U41541 Caenorhabdi
C 868	17	2.0	13833	3	AF070687	Drosophil	AF070687	C 941	17	2.0	40216	3	L16559	L16559 Caenorhabdi

942	17	2.0	41131	3	AF038609	AF038609	Caenorhab	SOURCE	Ehrlichia canis.
943	17	2.0	41135	9	AC090631	AC090631	Homo sapi	ORGANISM	Ehrlichia canis
944	17	2.0	41229	3	CEC15H1	Z81035	Caenorhabdi	REFERENCE	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
945	17	2.0	41324	3	AF043702	AF043702	Caenorhab	AUTHORS	Anaplasmataceae; Ehrlichia.
946	17	2.0	41609	3	AF016451	AF016451	Caenorhab	TITLE	1 (bases 1 to 6913)
947	17	2.0	42130	9	HSAC000355	AC000355	Human Cos		Ohashi,N., Unver,A., Zhi,N. and Rikihisa,Y.
948	17	2.0	42689	8	SPCC330	AL031603	S.pombe c		Cloning and characterization of multigenes encoding the
949	17	2.0	42874	9	AF181449	AF181449	Homo sapi		immunodominant 30-kilodalton major outer membrane proteins of
950	17	2.0	43907	3	AF078790	AF078790	Caenorhab	JOURNAL	Ehrlichia canis and application of the recombinant protein for
951	17	2.0	44000	6	AX252459	AX252459	Sequence	MEDLINE	serodiagnosis
952	17	2.0	44504	9	AC022089	AC022089	Homo sapi	REFERENCE	J. Clin. Microbiol. 36 (9), 2671-2680 (1998)
953	17	2.0	46588	9	AL135938	AL135938	Human DNA	PUBMED	98371112
954	17	2.0	47319	6	AX059533	AX059533	Sequence		9705412
955	17	2.0	49278	9	AL845557	AL845557	Human DNA		2 (bases 1 to 6913)
956	17	2.0	49355	2	AC130520	AC130520	Rattus no	REFERENCE	Ohashi,N., Rikihisa,Y. and Unver,A.
957	17	2.0	49925	9	AL844203	AL844203	Human DNA	AUTHORS	Analysis of transcriptionally active gene clusters of major outer
958	17	2.0	50016	9	AF274855	AF274855	Homo sapi	TITLE	membrane protein multigene family in Ehrlichia canis and E.
959	17	2.0	50302	2	AC095816	AC095816	Rattus no	JOURNAL	chaffeeensis
960	17	2.0	50362	2	AC091572	AC091572	Homo sapi	MEDLINE	Infect. Immun. 69 (4), 2083-2091 (2001)
961	17	2.0	51059	9	AL139417	AL139417	Human DNA	PUBMED	21153566
962	17	2.0	53135	9	AC116621	AC116621	Homo sapi	REFERENCE	1254561
963	17	2.0	53292	8	AB046432	AB046432	Arabidops	AUTHORS	3 (bases 1 to 6913)
964	17	2.0	53736	2	AC068452	AC068452	Mus muscu	TITLE	Ohashi,N., Rikihisa,Y. and Unver,A.
965	17	2.0	54492	9	AL590703	AL590703	Human DNA	JOURNAL	Direct Submission
966	17	2.0	54655	2	AC011215	AC011215	Homo sapi	FEATURES	Submitted (29-NOV-2000) Veterinary Biosciences, Ohio State
967	17	2.0	54705	2	AC006094	AC006094	Homo sapi	source	University, 1925 Coffey Rd., Columbus, OH 43210, USA
968	17	2.0	54940	9	AL596211	AL596211	Human DNA	gene	1. .6913
969	17	2.0	55008	2	AC023632	AC023632	3	CDS	/organism="Ehrlichia canis"
970	17	2.0	55037	2	AC009548	AC009548	Homo sapi		/strain="Oklahoma"
971	17	2.0	55256	9	HSJ735G18	AL096703	Human DNA		/db_xref="taxon:944"
972	17	2.0	55367	9	AC023305	AC023305	Homo sapi		329. .1420
973	17	2.0	55457	2	AC022835	AC022835	Homo sapi		/gene="purk"
974	17	2.0	55569	9	HS824J5	AL034552	Human DNA		/gene="purk"
975	17	2.0	55902	5	AF325922	AF325922	Takifugu		/codon_start=1
976	17	2.0	55965	9	AC104653	AC104653	Homo sapi		/transl_table=1
977	17	2.0	56403	9	HS4418064	AJ418064	Homo sapi		/product="phosphoribosylaminoimidazole carboxylase"
978	17	2.0	56511	3	AC084467	AC084467	Caenorhab		/protein_id="AAK31311.1"
979	17	2.0	57032	2	AC100894	AC100894	Mus muscu		/db_xref="GI:13591682"
980	17	2.0	57348	2	AC101362	AC101362	Mus muscu		SPSINVNTVILDNVYNNKELLLEFASVSDIATLEFENIPSTTIDILSQTKVYFGKT
981	17	2.0	57896	2	AC090335	AC090335	Homo sapi		translacion="MLHNSYVSGSTIGIIGGQGLKMWISIAATNLGYKTHLLTDNQD
982	17	2.0	57919	2	AC017856	AC017856	Drosophi		SPSINVNTVILDNVYNNKELLLEFASVSDIATLEFENIPSTTIDILSQTKVYFGKT
983	17	2.0	57991	9	AF013593	AF013593	Homo sapi		ALHISONREKQVIRNLGKTANFRIDNVNLVHNLVLEGLYPAKLLKTLGVDGKG
984	17	2.0	58270	9	AF250325	AF250325	Homo sapi		QYLITQPGDLNTLSTLNQOEYILLEKEVKHKEISIIISKSINGSIEFFPVAENHTN
985	17	2.0	58346	2	AC100846	AC100846	Homo sapi		GILATSSVPASISSEINQAKIALQIAESFNLVGLIAVEFFIDDEGLIWEIAPRP
986	17	2.0	58958	9	AC091813	AC091813	Mus muscu		HNHWSLDACNISQFQLPKPKVLLFPCTMONILGDDIYNYQYTKIN
987	17	2.0	59043	2	AC100881	AC100881	Human DNA		ESIYIYGTAKSKNRKMGHINRLKHEI"
988	17	2.0	59723	9	AL137803	AC092864	Homo sapi		1756. .2598
989	17	2.0	59892	9	AC092864	AC092864	Homo sapi		/gene="p30-10"
990	17	2.0	61417	2	AC023564	AC023564	Homo sapi		1756. .2598
991	17	2.0	61417	2	AC023564	AC023564	Homo sapi		/gene="p30-10"
992	17	2.0	61510	9	AL359436	AL359436	Human DNA		/note="P30 family"
993	17	2.0	61555	2	AC090334	AC090334	Homo sapi		/codon_start=1
994	17	2.0	61851	9	AC022240	AC022240	Homo sapi		/transl_table=1
995	17	2.0	62268	9	HS1178121	AL109852	Human DNA		/product="major outer membrane protein p30-10"
996	17	2.0	62423	8	AP004554	AP004554	Oryza eat		/protein_id="AAK31312.1"
997	17	2.0	62692	2	AC107846	AC107846	Mus muscu		/db_xref="GI:13591683"
998	17	2.0	62800	2	AC084216	AC084216	Homo sapi		translacion="MNYKILVRSALISLSILPYSQSFADPVGSRNDNKEGFYISAK
999	17	2.0	62825	2	AC090452	AC090452	Homo sapi		YNPSISHRFKFAEETPEINGNTSLTKVGLKKDGDITKKDDFTVAFGIDFQNNLIS
1000	17	2.0	62825	2	AC090452	AC090452	Homo sapi		FGSGSIGYMDGPRIEAAAYQQFNKNNTDNGEYKGFALSRDAMEDQQYVV
									LKNDGITFMSLMVNTCYDITAEQVSVFVACAGIADLITFKDLNLKFAVQKGIGIS
									YPIPEVSARFYGYYHGVIGNKFEKIPVITPVVLNDAPQTTSASVTLVDGVFGGEIGM
									RFT"
									2926. .3756
									/gene="p30-4"
									2926. .3756
									/gene="p30-4"
									/note="P30 family"
									/codon_start=1
									/transl_table=1
									/product="major outer membrane protein p30-4"
									/protein_id="AAK31313.1"
									/db_xref="GI:13591684"

## ALIGNMENTS

RESULT 1  
 AF324792  
 LOCUS  
 DEFINITION  
 AF324792  
 AF324792.1 GI:13591681  
 AF324792 6913 bp DNA linear BCT 11-APR-2001  
 Ehrlichia canis phosphoribosylaminoimidazole carboxylase (purK)  
 gene, complete cds; major outer membrane protein gene cluster 2,  
 complete sequence; and u6 gene, partial cds.  
 AF324792  
 AF324792.1 GI:13591681  
 KEYWORDS

```
/translation="NWCKKILITLVSLTLLPGISGSKPIHENNTGNFYIIGRVY
PSVISHGPNFSAKEKXNTGTGILGKESWTGGIILDKHAAFNTPNTSKFYENNPFLGF
AGVIGYSPIRESEVSEVTEVDVQNGDKNNDAAHKYKASNDSSKTMKSQGFVLK
NEGSLDISLWANCYDIIINRMPSPYTCAGITGTLFMEDANIKAAAYOGKGLKFNPF
ISPENISMGVHFHKVTNNNEPRVPVLLTAGGLAPDNLPAIVKLISICHEGLFEGYRVYSF
I
```

**gene**

3967. 4830

/gene="p30a"

3967.4830

/gene="p30a"

```
/note="P30 family"
```

```
/codon_start=1
/translation_table=11
```

[voicemail-combo/](#)

```

/product= "major outer membrane protein F30a
/protein id="AAK31314.1"

```

```
/db xref="GI:13591685"
```

```
/translation="MKYKKTFT
```

YQKLGKLTGNTINSRVSFAGGHFKVIGNEPKGIPTLPLDGSNIYQKQSAATVTLDVCH  
 AGATGYSISNARELEQVSHIEI PTDKNPNNYNNINNDTSKLYCALSHGSHICSDGNSGQWYT  
 AKDKFVLKNEGGLDVSFMLNACYITKKNPFSPIYACIGITDLISMETFTQNKIS  
 TASHFGISAKESQFTVGLDQRLSHNNIINNDDTSKLYQNSYFKYKNPFLGJ  
 AGATGYSISNARELEQVSHIEI PTDKNPNNYNNINNDTSKLYCALSHGSHICSDGNSGQWYT

gene

complement (5309. .&gt;6913)

`/gene="v6"`

```
complement(5309. >6913)
```

```
/gene="u6"
```

```
/codon_start=1
/translation_table=1
```

```
transl_table=1
```

```
/product="AKK31"
/protein id="AKK31"
```

```
/db xref="GI:13591686"
```

```
/translation="KLIVTQDKRV"
```

VLKDBYCDQNAVGQIQYKHQVDTKVCIDGINTIYGVLSKSHSVYDQKHTKLKPLVQD  
 MYRVLDTDKQYQIQADVEYDLTKNQNDLSDILINDENYVDFSEDEHNRKTLKPLVQD  
 DMPNCLKNSFVLYQNTKLTADNKSITLPEVFTQIQKDSLANSNVEFVPLNKQKIL  
 SAKDPMLEKTRFDTHVLFVNDNKEMLSDPKQLQQLDQVDSMGWCISDSFESTIYV  
 PEDMTESDKFDNALPVLQVNDKSESSVLLHYPKLSIYVPGKCDLTKVMEGLQSGBSG  
 DLTKISKIYKQKQGEQYMGHFTPKSDYKPVKLYFRKMENKSKSNDKHLKNDSDYD  
 GMDVATFEFCQKQERLHDGEVCEVQTANGGCKQLQHDTRKNVKNKNLNVMDKRLMYRNV  
 NVHSVRSRNSVPHDKRIRFIYQNPFIYTHALWEBEITEKDECKSGKLIAENNSKYFGAGCG  
 UNVSTKSTSTPBGASCMWIKTQKSDYDVRH

1149 C 888 g 2356 F

## ORIGIN

Query Match	100.0%;	Score 840;	DB 1;	Length 6913;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 840;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy 1 ATGAAATTATAAGAAAATTCCTAGTAAGAGCGGTTAATCTCATTAATGTCAATCTTTACCA 60

Db 1756 ATGAAATTATAAGAAAATTCCTAGTAAGAGCGGTTAATCTCATTAATGTCAATCTTTACCA 1815

```
/strain="Jake"
/db_xref="taxon:944"
<1..146
/gene="p28a"
<1..146
/gene="p28a"
/notes="outer membrane protein"
/codon_start=3
/transl_table=11
/product="p28"
/protein_id="AAG14355.1"
/db_xref="GI:10181083"
/translation="SIISHYSVKNLPVQVNTLEBYPRVTSIAITLDIGLGGEGIR
PIF"
170..1051
/gene="p28-1"
170..1051
/gene="p28-1"
/notes="outer membrane protein"
/codon_start=1
/transl_table=11
/product="p28-1"
/protein_id="AAG14356.1"
/db_xref="GI:10181084"
/translation="MNNKLKFTIINTVLVCLLSPNIISSKALNNNAKKYGLXISGO
YKPSVSVSNFSVKTNTVITKNLIALKKDVSIEYTKDASVGINSPSFTPIYTAVFQ
DNSVNFNGTIFAEGRVREIEGSEYEDVKNPGGYLSDAIRYFALAREMKNSFT
PKRVSIFHTWMDGLSIISIVNVCYDPSLNNLSISPIYCGAGVDAIEFDFVL
HIKAYQKGLIAYLPSNLSIFASLYHKVGNQFNKLNVQHVAAELASIPKITSAVA
TLNIGYFGGEIGARLTF"
1599..2441
/gene="p28-2"
1599..2441
/gene="p28-2"
/notes="outer membrane protein"
/codon_start=1
/transl_table=11
/product="p28-2"
/protein_id="AAG14357.1"
/db_xref="GI:10181085"
/translation="MNYKKILVRSALISLMSILPYQSPADVPVGSRTDNKKGFIYSK
YNPSISHRFKFAETPTNGTSLTKVFLGKKDGIITKKDDFTFRAVAGIDPQNNLIS
FGSGIGYSMDGPRIELEAAQYQPNKNTDNDNGEYVYKHFAIYKRDAMEDQOYVV
LKNDGITSMVNTCYDITAEQSVFVYACAGIGADLITIFKOLNLKFAVQKIGIS
YPTFEVSAFLIGYTHGVGNKFEKIPVITPVVLNDAPQTTSASVTLVDVGFPGGEIGM
RFTF"
2769..3599
/gene="p28-3"
2769..3599
/gene="p28-3"
/notes="outer membrane protein"
/codon_start=1
/transl_table=11
/product="p28-3"
/protein_id="AAG14358.1"
/db_xref="GI:10181086"
/translation="MNCCKILITTLVSLTLLPGISPSKPIHENTTGNFYIICKYV
PSIHFGFSAKEKNTTIGFLKESWTGGIILDKHAAFNIPNYSKYENNPFLGF
ACVIGYSYEFSEYETFDQNPQDKFNNDARHYCALSDNSDKTKMSKGFVFLK
NEGLSDISLMLNVCYDIINKRPFSPYICAGIGTDLIPMFDAINHKAAYQKGLGFNP
ISPENANISGVHFKVNTNNEFRVPLLTAGGLAPDNLFAIVKLSICHFGLEFGYRVSF
"
3810..4673
/gene="p28-4"
3810..4673
/gene="p28-4"
/notes="outer membrane protein"
/codon_start=1
/transl_table=11
/product="p28-4"
/protein_id="AAG14359.1"
/db_xref="GI:10181087"
/translation="WKYKKTFVTALVLLTSFTHPIFPYSPARASTIHNFIYSKGYM
TASHGIFSAKEQSTFKVLGLDQRLSHNIINNNTAKSLKVQNSYFKYKQNPFLGF
AGAIGYSIGNSRIELEVSHIEITDKNPGNNVLDNSHKYCALSHGSHICSDNSGDWYT
AKTDKFLVLLKNEGLLDVDFMLNACYDITTEKMPSPYICAGIGTDLISMFTTQNKIS
YQKGLGLNYTINSRVSVFAGGHFKHVGIGNEFKGIPTLPLPGSNIKVQQSATVLDVCH
FGLBISGRFF"
4970..5821
/gene="p28-5"
4970..5821
/gene="p28-5"
/notes="outer membrane protein"
/codon_start=1
/transl_table=11
/product="p28-5"
/protein_id="AAG14360.1"
/db_xref="GI:10181088"
/translation="MNCCKVFTISALISSIYFLPNVSYSNPVGNSWYGNFYISGKYM
PSVPHFGIFSAEEKKTIVYGLKENWAGDAISSQSPDNFTIRNYSFKYANKFLG
FAVAIGYSIGSPRIEVEVSEAFDVKNPQNDYKNGAYRICALSHQDDDDMTSATDK
FVYLINSEGLNISFMTNICYETASKNIPLSYICAGIGTDLIHMFTTHPKISYQKGL
GLAYFVSAESSVSFGIYFHKIINNKFKNVPAVPIINSDIIVGVQPQFATVTLNVCFGLE
LGCRNF"
6164..7006
/gene="p28-6"
6164..7006
/gene="p28-6"
/notes="outer membrane protein"
/codon_start=1
/transl_table=11
/product="p28-6"
/protein_id="AAG14361.1"
/db_xref="GI:10181089"
/translation="MNCCKILITTLALMLYAPSIISFSDTIQDNTGTSFYISGKYV
SVSHFGFSAKEERNSTVGVGLKHDWNGGTISNSPENFTVQNSYFKYENNPFLGF
AGAIGYSGGPRIEFLVYETFDVKNNYKNGAHRYCALSHSSATSMSANKFV
FLKNEGLDLSFMINACYDIIIEGMPSPYICAGIGTDLVVSMEFAVNPKISYQKGLG
GYSISSEASVFIGGHRVIGNEFRDIPAMVPSGSLPENQFATVTLNVCHFGLELGG
RNF"
7349..8185
/gene="p28-7"
7349..8185
/gene="p28-7"
/notes="outer membrane protein"
/codon_start=1
/transl_table=11
/product="p28-7"
/protein_id="AAC64550.2"
/db_xref="GI:10181082"
/translation="MNCCKILITTLALISLMYSIPSIISFSDTIQDNGNMGNFYISGKYV
PSVSHFGFSAKESKSTVGVGLKHDWDSIPILKNKHADFAVPNYSFRYENNPFLGF
AGAIGYSGGPRIEFEISYEAFDVKSNINYQNDARHYCALSHSSTAAMEADKVFLLK
NEGLDILAINACYDILINDKVPYSPYICAGIGTDLISMFEATSPKISYQKGLISVS
INPETSIVFIGGHFRIIGNEFRDIPAIVPSNSTTISGFPQFATVTLNVCHFGLELGGRF
NF"
8538..9404
/gene="p28-8"
8538..9404
/gene="p28-8"
/notes="outer membrane protein"
/codon_start=1
/transl_table=11
/product="p28-8"
/protein_id="AAG14362.1"
/db_xref="GI:10181090"
/translation="MNCCKRFFIASALISLMSFLPSVFSFSESIHEDNINGNFYISAKYM
PSAHFGFVSKEEKNTTGVGLKQWDGATIKDASSHTIDPSTIFSIISNYSFKYE
NNPFLGAGIYGSGGPRVEFEVSYEIFDVQNGSNYKNDARHYCALSRGTGMPQA
GHONKFPVLEKNEGLDLSLAINACYDITIDSMFPSPYICAGIGSDLSMFTTNPKIS
YQKGLGVSYISPEASVFGGHFHRVIGNEBPKDIPAITPACATBIKGTQFTVTLNIC
HFGLELGRFTF"
complement(10031..10846)
/gene="p28-9"
complement(10031..10846)
/gene="p28-9"
```

BASE COUNT		4065 a	1825 c	1662 g	3777 t
ORIGIN					
Query Match 100.0%; Score 840; DB 1; Length 11329;					
Best Local Similarity 100.0%; Pred. No. 0;					
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
Qy	1	ATGAATTATAGAAATTTCTAGTAAGAGCGGTTAATCTCAATTAATGCTCAATCTTACCA	60		
Db	1599	ATGAATTATAGAAATTTCTAGTAAGAGCGGTTAATCTCAATTAATGCTCAATCTTACCA	1658		
Qy	61	TATCAGTCTTTTGAGATCTCTGAGTTCAGAACTAATGATTAACAAGAGGCTTCTTAC	120		
Db	1659	TATCAGTCTTTTGAGATCTCTGAGTTCAGAACTAATGATTAACAAGAGGCTTCTTAC	1718		
Qy	121	ATTAGTCAAAAGTCAATCCAAAGTATATCAACATTTAGAAAATCTCTGCTGAAGAACT	180		
Db	1719	ATTAGTCAAAAGTCAATCCAAAGTATATCAACATTTAGAAAATCTCTGCTGAAGAACT	1778		
Qy	181	CCTATTATGAACAAATCTCTCAATTAAGAAAGTTTCGAGCTAAAGAAAGATGGTGAT	240		
Db	1779	CCTATTATGAACAAATCTCTCAATTAAGAAAGTTTCGAGCTAAAGAAAGATGGTGAT	1838		
Qy	241	ATAACAAAAAGAGCATTTTACAAGAGTACTCCAGGCATTGATTTTCAAAATAACTTA	300		
Db	1839	ATAACAAAAAGAGCATTTTACAAGAGTACTCCAGGCATTGATTTTCAAAATAACTTA	1898		
Qy	301	ATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGACGCGCAAGAAATAGAACTTGAA	360		
Db	1899	ATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGACGCGCAAGAAATAGAACTTGAA	1958		
Qy	361	GCTGCATATCACAATTTAATCCAAAAACCCGATATCAATGATGATGATTAATGGTGAA	420		
Db	1959	GCTGCATATCACAATTTAATCCAAAAACCCGATATCAATGATGATGATTAATGGTGAA	2018		
Qy	421	TACTATAAATTTTGCAATTTCTGTAAGATGCAATGGAAGATCAGCAATATGTAGTA	480		
Db	2019	TACTATAAATTTTGCAATTTCTGTAAGATGCAATGGAAGATCAGCAATATGTAGTA	2078		
Qy	481	CTTAAAAATGCGGCATAACTTTTATGTCAATGATGGTTAATCTGCTATGACATTACA	540		
Db	2079	CTTAAAAATGCGGCATAACTTTTATGTCAATGATGGTTAATCTGCTATGACATTACA	2138		
Qy	541	GCTGAAGAGTATCTTTCGTACCATATGCTATGCGAGTATAGGAGCAGATCTTATCACT	600		
Db	2139	GCTGAAGAGTATCTTTCGTACCATATGCTATGCGAGTATAGGAGCAGATCTTATCACT	2198		
Qy	601	ATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGCAAAATAGGTATTAGTTACCT	660		
Db	2199	ATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGCAAAATAGGTATTAGTTACCT	2258		
Qy	661	ATCACACAGAGTCTCTGCAATTTTATGGTGGATACCTACCATGGCGTTATTGGTAATAA	720		
Db	2259	ATCACACAGAGTCTCTGCAATTTTATGGTGGATACCTACCATGGCGTTATTGGTAATAA	2318		
Qy	721	TTTGAGAAGATACCTGTAATTAATCTCTGTAGTATTAATGATGCTCTCAAAACCAATCT	780		
Db	2319	TTTGAGAAGATACCTGTAATTAATCTCTGTAGTATTAATGATGCTCTCAAAACCAATCT	2378		
Qy	781	GCTTCAGTAACTCTTGAGTTGGATCTTTTGGCGGAGAAATTTGCAATGAGTTTCACTTC	840		
Db	2379	GCTTCAGTAACTCTTGAGTTGGATCTTTTGGCGGAGAAATTTGCAATGAGTTTCACTTC	2438		

/note="outer membrane protein"	
/codon_start=1	
/transl_table=11	
/product="p28-9"	
/protein_id="AAG14363.1"	
/db_xref="GI:10181091"	
/translation="MNKRFVVGVTLSFTVFVFLSDGAFSDANFSEGRRLYIGSQYKV	
GIPNFSNFBSETIPGIYTKKIPALGLDKSEINTNSNTRSYDTPYASSPAGSFGIIGY	
YVNDPRVFEGSYENFPERQWYSPENQYKFPALSRNATNSDNKFTVLENNVYDKS	
LVNVCVDIASGSIPLAPYMGAGVADYIKFLGISLPKFSYQVRFVGNVYNVNTMLF	
GGGYHKVVGDRHVERVEIAYHPTALSIVPRTTSATLNTDYFCWEIGFRPAL"	
RESULT 3	
AF078553	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
PUBMED	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
PUBMED	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REMARK	
COMMENT	
FEATURES	
source	
CDS	
gene	
CDS	

28254 bp DNA linear BCT 02-APR-2001  
Ehrlichia canis major outer membrane protein P30 multigene cluster  
1, complete sequence.  
AF078553 AF078554 AF078555 AH006958  
AF078553.2 GI:13512584  
Ehrlichia canis.  
Ehrlichia canis  
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
Anaplasmataceae; Ehrlichia.  
1 (bases 1 to 28254)  
Ohashi, N., Unver, A., Zhi, N. and Rikihisa, Y.  
Cloning and characterization of multigenes encoding the  
immunodominant 30-kilodalton major outer membrane proteins of  
Ehrlichia canis and application of the recombinant protein for  
serodiagnosis  
J. Clin. Microbiol. 36 (9), 2671-2680 (1998)  
98371112  
9705412  
2 (bases 1 to 28254)  
Ohashi, N., Rikihisa, Y. and Unver, A.  
Analysis of transcriptionally active gene clusters of major outer  
membrane protein multigene family in Ehrlichia canis and E.  
chaffeensis  
Infect. Immun. 69 (4), 2083-2091 (2001)  
21153566  
11254561  
3 (bases 1 to 28254).  
Ohashi, N., Unver, A., Zhi, N. and Rikihisa, Y.  
Direct Submission  
Submitted (16-JUL-1998) Department of Veterinary Biosciences, The  
Ohio State University, 1925 Coffey Road, Columbus, OH 43210, USA  
4 (bases 1 to 28254)  
Ohashi, N., Rikihisa, Y. and Unver, A.  
Direct Submission  
Submitted (29-NOV-2000) Department of Veterinary Biosciences, The  
Ohio State University, 1925 Coffey Road, Columbus, OH 43210, USA  
Sequence update by submitter  
On or before Apr 2, 2001 this sequence version replaced gi:3790556,  
gi:3790555, gi:3790558, gi:3790557.  
Location/Qualifiers  
1..28254  
/organism="Ehrlichia canis"  
/strain="Oklahoma"  
/db\_xref="taxon:944"  
<1..615  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical transcriptional regulator"  
/protein\_id="AAK28679.1"  
/db\_xref="GI:13512585"  
/translation="PEYANEIKAHDPDIEDLIKNIQHQKFTGEGIRLDVDDYASKNF  
KKGASTKELKATKVRPHPCVCKEIKRQIRMGSONOLANKLKITFOOVQKY  
EKTNRIVISRLYQLASVLNVEDRIMLKQEDLKNICNDNPITPPHALRDNEKFLP  
EPNDSKIDSKVELVMVRYATCIKNEKVRNIYNLVKALSLDNKS"  
843..1733  
/gene="p30-19"  
843..1733  
/gene="p30-19"  
/notes="P30 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein p30-19"  
/protein\_id="AAK28680.1"  
/db\_xref="GI:13512586"  
/translation="MKLLVHLDNIMIKFSAIGIVFSAIFALFAPNAPSPVPIDFSNES  
EMAGPASAQYINIGFRPSPISAKYTKDESKELTFLSKESTETIDLKKAGDFKKG  
YSPVNRNNTYFGSAGYGGGURVEGFTFRFDVKQKYNKPDGHRYPALSKDSEI  
QNSSSGSSNNKDYVMKNEGFNAISLMFNACYDMLIGNSSLVNPNACIGIGQGIIRFL

```

gene
2491. .3318
/ gene="p30-18"
CDS
2491. .3318
/ gene="p30-18"
/ note="p30 family member"
/ codon_start=1
/ transl_table=11
/ product="major outer membrane protein P30-18"
/ protein_id="AAK28681.1"
/ db_xref="GI:13512587"
/ translation="MNKLSLLYLIALILFTSHVSALVLDNHLNVYFGIQKPAHHL
SNLLKESKSDVEVLAKYDAIGFPLSTKEVNNFTIKNPHYDNNRGLGSVIFGYI
YKNNPRISEISHEIFQKNEGRKVGFEKFAKFPSPSTQGYRHVTLINNGISTT
SALINACDVLIPAHNIITYSLGFGIDIVDFLSKYTTKFSHQKLGASYPISHMSV
FTEVYHGLFGKFKPEQLPLNYNANTSPQPPHVTHTASAILSIGYGGVGIKFI"
3425. .3871
/ codon_start=1
/ transl_table=11
/ product="unknown function U2"
/ protein_id="AAK28682.1"
/ db_xref="GI:13512588"
/ translation="MNKVLQNEVLLALLPYSLHNSIDIEVNEOPIKLSRLNKNTI
KSQKHLAKYILLKIKISSIGWELFESNDYSNQVLYLRTSLSKLVSTIALDFDK
NIFLNKVLTSLRNFSANYFPSKHNITNCFVSRAYFELSCSY"
3882. .4832
/ gene="p30-17"
3882. .4832
/ gene="p30-17"
/ note="p30 family member"
/ codon_start=1
/ transl_table=11
/ product="major outer membrane protein P30-17"
/ protein_id="AAK28683.1"
/ db_xref="GI:13512589"
/ translation="MNYAKVFLVLCIIFLPBLSFATNNNVFVHEIGKSIGHFYIGVQ
YKGPHPNRSIADDSFTNLLAISHTKDYLFESYSTEVRGLFSPQEQNNLHYATGG
STTLTKDSDNPIPGYNTPTDNLGIVGGIYGSINNLRIELEAFYEFKNIKAPTGY
NYDTYFAIATVVKYTKPVHYHCHMNTGIIILSLFVNTCYDFTLKAKIAPYLCI
GVGGIDFPLGQTRUKASYQAKGLSYAISPNLTFVDSGFHGYMNNFPGLLDVYPT
DISVSPSGDNATAYSEFTTMLAKNLMIFLAGSIGIRIS"
4848. .5699
/ gene="p30-16"
4848. .5699
/ gene="p30-16"
/ note="p30 family member"
/ codon_start=1
/ transl_table=11
/ product="major outer membrane protein P30-16"
/ protein_id="AAK28684.1"
/ db_xref="GI:13512590"
/ translation="MYKLYYLSFIISLAQLFSGFAFSDIKNNNHGSIYTIKYOPTI
SNFNKFIKETDFDEDPIDFIAPNTNFPDLKHNYSVLXKDSYKFPYENDLSGL
ALSGLAKVNLRIEPEGYSKNPDTKRLAYHSRCHGKFFAIPRTSNFVGIPENDNTYV
AKNGGISIISNINLCSETKPKNFTPIQICIGIGDFEIEIPDMRVKFTYQKVGISYP
ITPKVLVLSQGHVHKVIGNKFKPLPIQVAPLQDFDNDSPEDKDVLTALLTLDLHFSE
IGLSFIF"
5716. .6510
/ gene="p30-15"
5716. .6510
/ gene="p30-15"
/ note="p30 family member"
/ codon_start=1
/ transl_table=11
/ product="major outer membrane protein P30-15"
/ protein_id="AAK28685.1"
/ db_xref="GI:13512591"
/ translation="MWINIYVYKTIKIALFLLPRVFSFSLIGNIEKSIKLLSVHIN
SOYKPSISQIENLYIOENSKSEKINILNSNNTIYNQLENSTNFRFIIGYFFKR
LRFAVEDYSIEHFKNDPSLKANLSKYKMYKNEDFNQFTIATDNKLSITSAINICY
DILINNTVLPHLCITAVGICSTGFENMRFKLLYQKIGLVILNSVWMLFFNVVYHK
VWRNKLKLLTQYSVDINAFDLDAITVLANTDIDIGFSGSEVGVRFFIN"

```

```

gene
6525. .7394
/ gene="p30-14"
6525. .7394
/ gene="p30-14"
/ note="p30 family member"
/ codon_start=1
/ transl_table=11
/ product="major outer membrane protein P30-14"
/ protein_id="AAK28686.1"
/ db_xref="GI:13512592"
/ translation="MLQRLNFINIILAFLLLPFQSFLLYIHDHEITONVGLYISSQ
YKPSIPYKFNPLEIENSHKTVELMGLANDVHTVEYVLKONTKFTNTPYSKAFRNSLIN
LSGAIGVYSGQRELEHSGVENFDPVASCNKNVKNARVIALVRDKPKGNIVQDHS
HSNMSYTFPKNGGISILSWINGCYDIAFSNVKLSYPVCAGIGGDFITLIPETWHIF
KYQCKFGISVLVSPSISIFANGHYHKVMDNVFNKLHVKIYIKQDADAPITTSARAKURI
GYSGSEVGVRFF"
7419. .8255
/ gene="p30-13"
7419. .8255
/ gene="p30-13"
/ note="p30 family member"
/ codon_start=1
/ transl_table=11
/ product="major outer membrane protein P30-13"
/ protein_id="AAK28687.1"
/ db_xref="GI:13512593"
/ translation="MNNKSLILGITILLSFSSLPKAFSVINHSDISSNVNGLYFTG
QYRPVSHFSGFTVRETNITATQQLVSLNTNKNENHIIERTNFSGIYTAQFDNAASF
GAIGSYPEGLKFEIEISYKFGKSTKNTQSTNAVIFALARQTTSNPNDKTVVMK
NSGLSVASVMINGNYSFNLVVSYPYICAGIGEDFIEFFDTLYIKLAYQCKLGVNS
LSSRFNIFADMYHKVIGNQFKNLVIHAVALDTFPKVTSAIATLNAVFGGEGVIRF
IL"
8266. .8823
/ codon_start=1
/ transl_table=11
/ product="unknown function U3"
/ protein_id="AAK28688.1"
/ db_xref="GI:13512594"
/ translation="MATYMIHAKIIEPMSPSILTYIILSLINNVDHNSPFIICIT
KLYKLILSSITKLLINVDKNIALHNILNYTLNKKSKFKSTQSFYEMFYTKIKOR
ILNPNISYFKYNTVTKNSNESYKSTYKSHNPNFNATYTTSTFHOHKLTYKYSITNQH
NKSTYKSPSKILAKNILLNNQNSYE"
8816. .9685
/ gene="p30-12"
8816. .9685
/ gene="p30-12"
/ note="p30 family member"
/ codon_start=1
/ transl_table=11
/ product="major outer membrane protein P30-12"
/ protein_id="AAK28689.1"
/ db_xref="GI:13512595"
/ translation="MNSKTFSLIGSILICLAACLPQSFSESSNVYTNKHTGLYIS
GLYKPSYSHFSDFSIKETTYNTALFGLKQDIISSILNKETTYNNNFVYPTAKFQD
DFASFTAVGIANNPRIIEGSEYEEFDVKNPGNYYTIDAHRYIALAREKTSYLS
PKSNFYIKNNGISIVILINGCYDILSNDKSVSPYICTGFGGDFIEFESAIPEKFA
YQKIGISYLSNIIILFTDGYTHKVINSQFKNLNVEHVVELLTDPAKVISATAFNLI
EYFGGEFGLFIF"

```

```

Query Match      100.0%; Score 840; DB 1; Length 28254;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAATTATAAGAAAAATCTAGTAAGAGCGCGGTAATCTCATTTAATGTCTTAATCTTACCA 60
Db 15562 ATGAATTATAAGAAAAATCTAGTAAGAGCGCGGTAATCTCATTTAATGTCTTAATCTTACCA 15621

Qy 61 TATCAGTCTTTTTCAGATCCTGTAGGTTCAAGACTAATGATACAAAGAGGCTTCTAC 120
Db 15622 TATCAGTCTTTTTCAGATCCTGTAGGTTCAAGACTAATGATACAAAGAGGCTTCTTAC 15681

Qy 121 ATTAGTCAAAAGTACAATCCAAGTATATCACATTTTAGAAAAATCTCTGCTGGAAGAACT 180
Db 15683 ATTAGTCAAAAGTACAATCCAAGTATATCACATTTTAGAAAAATCTCTGCTGGAAGAACT 180

```

Db 15682 ATTAGTCGAAGTACAAATCCAAAGTATATACACATTTAGAAAAATCTCTGCTGAAGAACT 15741

Qy 181 CCTATTAAATGAACAATAATCTCTCACTAAAAAGTTTTCCGACTAAAGAAAGATGGTGAT 240

Db 15742 CCTATTAAATGAACAATAATCTCTCACTAAAAAGTTTTCCGACTAAAGAAAGATGGTGAT 15801

Qy 241 ATACAAAAAAGACGATTTTACAGAGTAGTCTCCAGGCATTTGATTTTCAAAAATCACTTA 300

Db 15802 ATACAAAAAAGACGATTTTACAGAGTAGTCTCCAGGCATTTGATTTTCAAAAATCACTTA 15861

Qy 301 ATATCAGGATTTTACAGGAAGTATTGGTTACTCTATGAGCGACCAAGAAATAGAACTTCAA 360

Db 15862 ATATCAGGATTTTACAGGAAGTATTGGTTACTCTATGAGCGACCAAGAAATAGAACTTCAA 15921

Qy 361 GCTGCATATCAACAATTTTAATCCAAAAACCGATAAACAATGATCTGATTAATGGTGAA 420

Db 15922 GCTGCATATCAACAATTTTAATCCAAAAACCGATAAACAATGATCTGATTAATGGTGAA 15981

Qy 421 TACTATAAACAATTTTGCATTTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTAGTA 480

Db 15982 TACTATAAACAATTTTGCATTTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTAGTA 16041

Qy 481 CTTAAAAATACGGCATAACTTTTATGTCATTGATGTTTAACTATGCTATGACATTAACA 540

Db 16042 CTTAAAAATACGGCATAACTTTTATGTCATTGATGTTTAACTATGCTATGACATTAACA 16101

Qy 541 GCTGAAGCAGTATCTTTTCGTACCATATGATGTCAGGATATAGGACGAGATCTTATCACT 600

Db 16102 GCTGAAGCAGTATCTTTTCGTACCATATGATGTCAGGATATAGGACGAGATCTTATCACT 16161

Qy 601 ATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGAAAAAATAGGTATTAGTTACCT 660

Db 16162 ATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGAAAAAATAGGTATTAGTTACCT 16221

Qy 661 ATCACACAGAGTCTCTGCATTTATGTCGTGATCTACCTACCTACCTGCTTATGTTAAATAA 720

Db 16222 ATCACACAGAGTCTCTGCATTTATGTCGTGATCTACCTACCTACCTGCTTATGTTAAATAA 16281

Qy 721 TTTGAGAAGATACCTGTAAATAACTCTCTGTAGTATTAATGATGCTCTCAACACACATCT 780

Db 16282 TTTGAGAAGATACCTGTAAATAACTCTCTGTAGTATTAATGATGCTCTCAACACATCT 16341

Qy 781 GCTTCAGTAACTCTGAGTTGGATCTCTTGGCGGAGAAATGGAATGAGTTTACCTTC 840

Db 16342 GCTTCAGTAACTCTGAGTTGGATCTCTTGGCGGAGAAATGGAATGAGTTTACCTTC 16401

RESULT 4

ECU72291 27190 bp DNA linear BCT 02-APR-2001

DEFINITION Ehrlichia chaffeensis strain Arkansas major outer membrane protein

ACCESSION U72291 AF021338

VERSION U72291.2

KEYWORDS GI:13511827

SOURCE Ehrlichia chaffeensis.

ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Anaplasmataceae; Ehrlichia

REFERENCE 1 (bases 14844 to 21136; 21479 to 22234)

AUTHORS Ohashi, N., Zhi, N., Zhang, Y. and Rikihisa, Y.

TITLE Immunodominant major outer membrane proteins of Ehrlichia chaffeensis are encoded by a polymorphic multigene family

JOURNAL Infect. Immun. 66 (1), 132-139 (1998)

MEDLINE 98084465

PUBMED 9423849

REFERENCE 2 (bases 1 to 27190)

AUTHORS Ohashi, N., Rikihisa, Y. and Unver, A.

TITLE Analysis of transcriptionally active gene clusters of major outer membrane protein multigene family in Ehrlichia canis and E. chaffeensis

JOURNAL Infect. Immun. 69 (4), 2083-2091 (2001)

MEDLINE 21153566

11254561

3 (bases 1 to 27190)

REFERENCE Ohashi, N., Zhi, N., Zhang, Y. and Rikihisa, Y.

AUTHORS Direct Submission

TITLE Submitted (24-SEP-1996) Department of Veterinary Biosciences, The Ohio State University, 1925 Coffey Rd., Columbus, OH 43210-1093, USA

JOURNAL

4 (bases 1 to 27190)

REFERENCE Ohashi, N., Zhi, N., Zhang, Y. and Rikihisa, Y.

AUTHORS Direct Submission

TITLE Submitted (27-AUG-1997) Department of Veterinary Biosciences, The Ohio State University, 1925 Coffey Rd., Columbus, OH 43210-1093, USA

JOURNAL

5 (bases 1 to 27190)

REFERENCE Ohashi, N., Rikihisa, Y. and Unver, A.

AUTHORS Direct Submission

TITLE Submitted (29-NOV-2000) Department of Veterinary Biosciences, The Ohio State University, 1925 Coffey Rd., Columbus, OH 43210-1093, USA

JOURNAL

REMARK Sequence update by submitter

COMMENT On or before Apr 2, 2001 this sequence version replaced gi:2853584, gi:2853273.

FEATURES

source

1..27190

/location=Qualifiers

/organism="Ehrlichia chaffeensis"

/strain="Arkansas"

/db\_xref="taxon:945"

315..956

/codon\_start=1

/transl\_table=1

/product="hypothetical transcriptional regulator"

/protein\_id="AAC02808.2"

/db\_xref="GI:13511828"

/translation="MSTHAKINKMEDTDEIKAHDLIAGLIDKNKQKFTKNGVNDDTSKNTKEGVSTKDLKNTTKIRPHVDECVGKEIKRQIRMGNSQNLANKLGITFOVKYKGNRIVISRLYQLASVNVEDRIMLKQEDLKNISCDHTVASTSTLKDNEKPIPFHDSKIDSKEVLMVMVRAITCIKNEKRNIIYNLVKALSLDNQ"

1136..2032

/gene="omp-1M"

1136..2032

/gene="omp-1M"

/note="omp-1 family member"

/codon\_start=1

/transl\_table=1

/product="major outer membrane protein omp-1M"

/protein\_id="AAK28658.1"

/db\_xref="GI:13511829"

/translation="MILINMKLFVHLDNIMTKFSAIGIVLSLVTLPACNVFASIPIDFSNESETAGYAGHYNIQPRFSPISVKYKTDTEKELTFLTKESTDPTTFKQKSEFNDKKGYSIYRNYTGFSGAIGYSGGLRVELEGATFVDVKQYKKNYRIFALCKKDSIBSTNSNGNHVMKNGEPRVSLTFNACYDMVSNLSLVPACIGIGQITNFLAGTNTHTLPKAKLIGLGLISPKTVIFANGYVVKTKDNTFNLVQYPLEKEAPKHIDPIACFNADNYGGEVGLRFL"

2498..3352

/gene="omp-1N"

2498..3352

/gene="omp-1N"

/note="omp-1 family member"

/codon\_start=1

/transl\_table=1

/product="major outer membrane protein omp-1N"

/protein\_id="AAK28659.1"

/db\_xref="GI:13511830"

/translation="MSKRSNRKFLVWMLILFTPHISLASVLDHNSMYGIVQVKPAROHLKLLIKESAANTVEVFGKKDLDLLTGKIDNTFNKIKNPYNNRNLGSGIFGYTNKNFRIESELSYETFKKNGYKRIKCEKHFALAKISGGSNNPANNKYVTLINNGISLTALINVCYDVGDKHNIITSCGLGVDITDLSKYTTTFYQGLGASVTVSPQSVFIEGYTHGLFGKFEKIPVNPYPCDPSPTPPNSKPHVHTTALAMLSIGYVGSIGIKFTL"

3459..3905

/codon\_start=1

/transl\_table=1

/product="unknown function UN2"

```

/protein_id="AAK28660.1"
/db_xref="GI:13511831"
/translation="MNNKSRGQVLLFAAYSIKFKYINCTSKFRLSKVCKFYVNPCLKPY
YKSTFLKGFVNYNYSKTSNLYHSLFKLFGIIRTNFKNLSVNFPTNPRVLIY
YKILKNKNSISINLLNKYINPTATSFINSVIIGIYVGFVNYN"
3915..4808
/gene="omp-1Q"
3915..4808
/gene="omp-1Q"
/notes="omp-1 family member"
/codon_start=1
/transl_table=11
/product="major outer membrane protein OMP-1Q"
/protein_id="AAK28661.1"
/db_xref="GI:13511832"
/translation="MSYAKVFILICLILLVPSLGFAYVNNDFLKDNIHGHYIGQYK
GVPRNRLVTNNIRLMSDECRSTIPHMVQSVQAQCTLPPEALELAKGLLHG
LFTLPYNTYKQNLGAGVIGYSTTHFRVEAFYEKNTLAPAGVLHKNFYEYFA
LATMDTKHQSAEDKYVMKNTGITLSPELINACVDFILKTRNVAPVLCVGGN
LFDLDQVSPFAQVQKVGISYFVSPNIAFFIDGSHGLHNNQFSDLPVDYSSGGP
TISAKNFANFLTSSIGIRFIS"
4825..5682
/gene="omp-1P"
4825..5682
/gene="omp-1P"
/notes="omp-1 family member"
/codon_start=1
/transl_table=11
/product="major outer membrane protein OMP-1P"
/protein_id="AAK28662.1"
/db_xref="GI:13511833"
/translation="MQKLYIFILISGLLLPKYVFCMQNNNDIGSYTIKQLTTPH
FKNFYKEDTDQIPLAKITANTKFDLKENFSPFHQTDYSKYQNDLIGL
SVGLFVKRFEFEAGYKNTKRLARYKSDGYKFAIPRKSEHGLDNTFGYTVAK
NNGSIINILNSETKYGKAFYPIYICIGVGDFIEIFDVMRIKFAVQCKYGVSPIT
SKLILSINGQYHKVIGNKPELLPVYQPVKELKRLVNTKSKDIDQDVTASLTNLEHFS
SEIGLSFIY"
5695..6513
/gene="omp-1T"
5695..6513
/gene="omp-1T"
/notes="omp-1 family member"
/codon_start=1
/transl_table=11
/product="major outer membrane protein OMP-1T"
/protein_id="AAK28663.1"
/db_xref="GI:13511834"
/translation="WMYMKKHCVIVTVITLFLPLLPIESLSALIGNVEKDLKVS
TVVSQYKPSIFPHFNFSIQSHPKKSSPEPKIKANLNILKSNAYNLQODNTSP
SGTIGYFKGLRLEAGCYQEPNVKNNSLIISNKHYSRIHDBENYALITNNKLSIA
SIMVNTCYDISINNTSIVPYLCTGIGEDLVGLFNTLHFKLAYQGVKMSYLNINILL
FSDIYHKVMGNRKNLYQVYVADPNISEETIPILAKLDIGYFGEIGIRFMFN"
6529..7416
/gene="omp-1U"
6529..7416
/gene="omp-1U"
/notes="omp-1 family member"
/codon_start=1
/transl_table=11
/product="major outer membrane protein OMP-1U"
/protein_id="AAK28664.1"
/db_xref="GI:13511835"
/translation="MTKRFNVNVLITLLEFLPLKSFYTTANNNTITQKVLGISGQ
YKPSIPHPKNSVEENDKVDLIGLTDVYITEHLRNDTKFNTHYIAKFNKNIFN
SSAIGYQGRPLYSYSGDGVVNYKQVADVNRVYFALVREKNGSNFSPKHET
SOPSDNPKSPYTLKNGGVFVAVIINGCYDFSNNTTISPYVCIGVGDFIEFFE
VMHIFACQSKVIGISYSPISPTIFADAHYHKVINNNKNNLHVKSYSKNSPTTISA
TAKLNIETPGEVGNRF"
7438..8277
/gene="omp-1V"
7438..8277
/gene="omp-1V"
/notes="omp-1 family member"

```

```

/codon_start=1
/transl_table=11
/product="major outer membrane protein OMP-1V"
/protein_id="AAK28665.1"
/db_xref="GI:13511836"
/translation="MSKKFTITIGTVLASSLSFISFSAIHNHHTGNNTSGIYITG
QYRPGVSHFNFSVKEINVDITQLVGYKKSASSIDPNTYGNFQGYTVTFODNAASF
KAGYSYPSIRLELEGSYEKFDVKDPKYSADKAFEPALARENTSTVPDAKYTVM
KNGLSVASIMINGCYDLSFNNLVSPYICAGIGEDFIEFDTHIKLAYQCKLGISY
YFFPKINVFAGGYHVRVIGNKFNLVNHNVTLDDEFPKATSAVALTNVAYVGGAGVK
FTF"
8300..8890
/codon_start=1
/transl_table=11
/product="unknown function UN3"
/protein_id="AAK28666.1"
/db_xref="GI:13511837"
/translation="MIYKELTRVGEYILAYLSFILTSTYIFLVNIRVNSLAICVI
SLRTNIFNVSTKLIKDKCRDTKFSNMNCLYKGLNLOIFYGIFSGFIRNFONTLI
IPNDSKCGFTLWDNPAHYTYTLTSGEYRNFEDILYENICQCKLLINNRSLNQ
HNKNTLVIIPIPNAREFSNEIRVNISINKSESYEC"
8880..9731
/gene="omp-1W"
8880..9731
/gene="omp-1W"
/notes="omp-1 family member"
/codon_start=1
/transl_table=11
/product="major outer membrane protein OMP-1W"
/protein_id="AAK28667.1"

```

Query Match 3.6%; Score 30; DB 1; Length 27190;  
Best Local Similarity 100.0%; Pred. No. 9e-05;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 523 ACTTGCTATGACATTACAGCTGAAGGAGTA 552  
|||||  
DB 16201 ACTTGCTATGACATTACAGCTGAAGGAGTA 16230

RESULT 5  
AF319940

LOCUS AF319940 2362 bp DNA linear BCT 02-APR-2002  
DEFINITION Cowdria ruminantium isolate Senegal major antigenic protein 1-2  
(mapi-2) and major antigenic protein 1-1 (mapi-1) genes, partial  
cde.

ACCESSION AF319940  
VERSION AF319940.1 GI:17224941

KEYWORDS  
Ehrlichia ruminantium.  
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
Anaplasmataceae; Ehrlichia.

SOURCE  
ORGANISM  
Bekker, C. P., Bell-Sakyi, L., Paxton, E. A., Martinez, D., Bensaïd, A.  
and Jongejans, F.  
1 (bases 1 to 2362)  
REFERENCE  
AUTHORS  
Bekker, C. P., Bell-Sakyi, L., Paxton, E. A., Martinez, D., Bensaïd, A.  
and Jongejans, F.  
Transcriptional analysis of the major antigenic protein 1 multigene  
family of Cowdria ruminantium  
JOURNAL Gene 285 (1-2), 193-201 (2002)  
MEDLINE 22035368  
PUBMED 12039046

REFERENCE 2 (bases 1 to 2362)  
Bekker, C. P., Paxton, E. A., Bell-Sakyi, L., Martinez, D., Bensaïd, A.  
and Jongejans, F.  
Direct Submission  
TITLE Submitted (07-NOV-2000) Bacteriology, Faculty of Veterinary  
JOURNAL Medicine, Yalelaan 1, Utrecht 3584CL, The Netherlands  
FEATURES  
source  
1..2362  
/organism="Ehrlichia ruminantium"  
/isolate="Senegal"  
/db\_xref="taxon:779"  
<1..228

gene



```

CDS
/gene="map1-2"
<1..228
/gene="map1-2"
/codon_start=1
/transl_table=11
/product="major antigenic protein 1-2"
/protein_id="AAL37179.1"
/db_xref="GI:17224942"
translation="KLAYQKGLGHSYPMSSNIKLFADGYYHKVTDNKFQNLKVIHVAD
LYNPLVTSATLNLIEYFGGEIGIRFGLK"
1613..>2362
/gene="map1-1"
1613..>2362
/gene="map1-1"
/codon_start=1
/transl_table=11
/product="major antigenic protein 1-1"
/protein_id="AAL37180.1"
/db_xref="GI:17224943"
translation="MNYKKILVRSALISLMSFLPYQSFAPVSSNNIGNENAKGFFYI
SAKNPSIPHRKFSABETPVYKDSPTKKVFGLKDGSIKYSDFTRTDISFEGQNN
FISFGSGSIGYIMDGRVIEAAQKFNPKNPANETDTSYKYHGLSRAETMTDKKY
VLTNNGVTFSLLMFNACDYDITAEGVFFIPYACAGIGADLISIFDDINLKFAYQKGI
ISVPTPEISAFIGGYHGVGNKVPKLV"
2 others
BASE COUNT 939 a 295 c 263 g 863 t
ORIGIN
Query Match 2.7%; Score 23; DB 1; Length 2362;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 TTACCATATCAGTCTTTTGCGA 77
|||||
Db 1667 TTACCATATCAGTCTTTTGCGA 1689

RESULT 6
AF125276 3507 bp DNA linear BCT 20-APR-1999
LOCUS Cowdria ruminantium isolate Highway major antigenic protein 1 like
DEFINITION protein and major antigenic protein 1 (MAP1) genes, complete cds.
ACCESSION AF125276
VERSION AF125276.1 GI:4589100
KEYWORDS Ehrlichia ruminantium.
SOURCE Ehrlichia ruminantium.
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
REFERENCE 1 (bases 1 to 3507)
AUTHORS Sulsona,C.R., Mahan,S.M. and Barbet,A.F.
TITLE The map1 Gene of Cowdria ruminantium Is a Member of a Multigene
FAMILY Containing Both Conserved and Variable Genes
JOURNAL Biochem. Biophys. Res. Commun. 257 (2), 300-305 (1999)
REFERENCE 2 (bases 1 to 3507)
AUTHORS Sulsona,C.R., Mahan,S.M. and Barbet,A.F.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-1999) College of Veterinary Medicine Department
of Pathobiology, University of Florida, P.O. Box 110880,
Gainesville, FL 32610, USA
FEATURES
source
Location/Qualifiers
1..3507
/organism="Ehrlichia ruminantium"
/isolate="Highway"
/db_xref="taxon:779"
948..1796
/notes="ORF2"
/codon_start=1
/transl_table=11
/product="major antigenic protein 1 like protein"
/protein_id="AAD26347.1"
/db_xref="GI:4589101"
translation="MNYKKILVRSALISLMSFLPYQSFAPVSSNNIGNENAKGFFYI
SAKNPSIPHRKFSABETPVYKDSPTKKVFGLKDGSIKYSDFTRTDISFEGQNN

```

```

FISFGSGSIGYIMDGRVIEAAQKFNPKNPANETDTSYKYHGLSRAETMTDKKY
VLTNNGVTFSLLMFNACDYDITAEGVFFIPYACAGIGADLISIFDDINLKFAYQKGI
ISVPTPEISAFIGGYHGVGNKVPKLVTLTDAPOQSTASVTLDAFGGEL
GVRFTF"
2155..3018
/gene="MAP1"
2155..3018
/gene="MAP1"
/codon_start=1
/transl_table=11
/product="major antigenic protein 1"
/protein_id="AAD26348.1"
/db_xref="GI:4589102"
translation="MNCKKIFITSLISLVSFLPGVSFSDVIQEDSNPAGSVYISAKY
MPTASHFGKMSIKEDSKNTQVFLGKKMDGVKTPSDSNTNSTIFTKDYSPFYENN
PFLTGAIGYSMNGPRIEVEYETPDVKNPGNYKNAHMYCALDTAAQNSNNGAG
LTTVMVKNENLTNLSIMLNACVDIMLDGIPVSPYVACAGITDLVSVINATNPKLSQ
GKLGISYSINSEASIFIGGHFRHVRVIGNEFKDIALTKIFTSKTGISNPGFASATLDVCH
FGIEIGRFFV"
BASE COUNT 1288 a 473 c 459 g 1287 t
ORIGIN
Query Match 2.7%; Score 23; DB 1; Length 3507;
Best Local Similarity 100.0%; Pred. No. 0.66;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 TTACCATATCAGTCTTTTGCGA 77
|||||
Db 1002 TTACCATATCAGTCTTTTGCGA 1024

RESULT 7
AF125274 3535 bp DNA linear BCT 20-APR-1999
LOCUS Cowdria ruminantium isolate Welgevonden major antigenic protein 1
DEFINITION like protein and major antigenic protein 1 (MAP1) genes, complete
cds.
ACCESSION AF125274
VERSION AF125274.1 GI:4589094
KEYWORDS Ehrlichia ruminantium.
SOURCE Ehrlichia ruminantium.
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
REFERENCE 1 (bases 1 to 3535)
AUTHORS Sulsona,C.R., Mahan,S.M. and Barbet,A.F.
TITLE The map1 Gene of Cowdria ruminantium Is a Member of a Multigene
FAMILY Containing Both Conserved and Variable Genes
JOURNAL Biochem. Biophys. Res. Commun. 257 (2), 300-305 (1999)
REFERENCE 2 (bases 1 to 3535)
AUTHORS Sulsona,C.R., Mahan,S.M. and Barbet,A.F.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-1999) College of Veterinary Medicine Department
of Pathobiology, University of Florida, P.O. Box 110880,
Gainesville, FL 32610, USA
FEATURES
source
Location/Qualifiers
1..3535
/organism="Ehrlichia ruminantium"
/isolate="Welgevonden"
/db_xref="taxon:779"
987..1835
/notes="ORF2"
/codon_start=1
/transl_table=11
/product="major antigenic protein 1 like protein"
/protein_id="AAD26343.1"
/db_xref="GI:4589095"
translation="MNYKKILVRSALISLMSFLPYQSFAPVSSNNIGNENAKGFFYI
SAKNPSIPHRKFSABETPVYKDSPTKKVFGLKDGSIKYSDFTRTDISFEGQNN
FISFGSGSIGYIMDGRVIEAAQKFNPKNPANETDTSYKYHGLSRAENADKKY
VLTNNGVTFSLLMFNACDYDITAEGVFFIPYACAGIGADLISIFDDINLKFAYQKGI
ISVPTPEISAFIGGYHGVGNKVPKLVTLTDAPOQSTASVTLDAFGGEL
GVRFTF"

```

```

gene      2210..3055
/ gene="MAPI"
CDS      2210..3055
/ gene="MAPI"
/codon_start=1
/product="major antigenic protein 1"
/protein_id="AAD26344.1"
/db_xref="GI:4589096"
/translation="MNCKKIFITSTLISLVSFLPGVSFSDAIQEDSSPVGVSYSYSAKY
MPTASHFGKMSIKEDSRDVKVFGKDKDGVKTSSNTIFTEKDYSPKYENNPLFGF
AGAIGYSNMGPRIEFEISYETFDVKNPGNYKNDAMVYCALDTSSTSGGAASVVMV
KNEULTDLSMLNACYDMLDGMVSPVYVACAGICTDLVSVNSTPNKLSYOGKLGISY
SINPEASIFIGGHFRVIGNEFKDITTSKIFNTSGATGPGFASATLVDVCHFGIEIG
GRFVF"
BASE COUNT 1290 a 445 c 480 g 1320 t
ORIGIN
Query Match      2.7%; Score 23; DB 1; Length 3535;
Best Local Similarity 100.0%; Pred. No. 0.66;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      55 TTACCATATCAGTCTTTTGCGA 77
|||||
DB      1041 TTACCATATCAGTCTTTTGCGA 1063

RESULT 8
AF125277      3538 bp DNA linear BCT 20-APR-1999
LOCUS      Cowdria ruminantium isolate Lemco3 major antigenic protein 1 like
DEFINITION      protein and major antigenic protein 1 (MAPI) genes, complete cds.
ACCESSION      AF125277
VERSION      AF125277.1 GI:4589103
KEYWORDS
SOURCE      Ehrlichia ruminantium.
ORGANISM      Ehrlichia ruminantium.
REFERENCE      1 (bases 1 to 3538)
AUTHORS      Sulsona,C.R., Mahan,S.M. and Barbet,A.F.
TITLE      The map1 Gene of Cowdria ruminantium Is a Member of a Multigene
          Family Containing Both Conserved and Variable Genes
JOURNAL      Biochem. Biophys. Res. Commun. 257 (2), 300-305 (1999)
REFERENCE      2 (bases 1 to 3538)
AUTHORS      Sulsona,C.R., Mahan,S.M. and Barbet,A.F.
TITLE      Direct Submission
JOURNAL      Submitted (02-FEB-1999) College of Veterinary Medicine Department
          of Pathobiology, University of Florida, P.O. Box 110880,
          Gainesville, FL 32610, USA
FEATURES
source      Location/Qualifiers
1..3538
/organism="Ehrlichia ruminantium"
/isolate="Lemco3"
/db_xref="taxon:779"
991..1839
/note="ORF2"
/codon_start=1
/transl_table=1
/product="major antigenic protein 1 like protein"
/protein_id="AAD26349.1"
/db_xref="GI:4589104"
/translation="MNYKKILVRSALISLMSFLPQSFAPVSSNNIGNENAKGFGYI
SAKNYSPHPRKFSABETPVYKDSPTKKVFLGKKGSIYKSDFTRTDISFEGQNN
FISGSGIGYIMDGPVEIAAYQKFNPKNPANETDTSYKYHGLSRAAMADKKY
VLTNNGVTFSLMFNACVDITAEGVPIPVACAGIGADLISIFDDINLKFAYOQKIG
ISYPTTPEISAFIGGYHGVGNKYNKVPVKLPVLTDPQSTASVTLDAQYFGGEL
GVRFTF"
2213..3058
/ gene="MAPI"
CDS      2213..3058
/ gene="MAPI"
/codon_start=1

```

```

/transl_table=1
/product="major antigenic protein 1"
/protein_id="AAD26350.1"
/db_xref="GI:4589105"
/translation="MNCKKIFITSTLISLVSFLPGVSFSDAIQEDSSPVGVSYSYSAKY
MPTASHFGKMSIKEDSRDVKVFGKDKDGVKTSSNTIFTEKDYSPKYENNPLFGF
AGAIGYSNMGPRIEFEISYETFDVKNPGNYKNDAMVYCALDTSSTSGGAASVVMV
KNEULTDLSMLNACYDMLDGMVSPVYVACAGICTDLVSVNSTPNKLSYOGKLGISY
SINPEASIFIGGHFRVIGNEFKDITTSKIFNTSGATGPGFASATLVDVCHFGIEIG
GRFVF"
BASE COUNT 1292 a 449 c 478 g 1319 t
ORIGIN
Query Match      2.7%; Score 23; DB 1; Length 3538;
Best Local Similarity 100.0%; Pred. No. 0.66;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      55 TTACCATATCAGTCTTTTGCGA 77
|||||
DB      1045 TTACCATATCAGTCTTTTGCGA 1067

RESULT 9
AF125279      3541 bp DNA linear BCT 20-APR-1999
LOCUS      Cowdria ruminantium isolate Antigua major antigenic protein 1 like
DEFINITION      protein and major antigenic protein 1 (MAPI) genes, complete cds.
ACCESSION      AF125279
VERSION      AF125279.1 GI:4589109
KEYWORDS
SOURCE      Ehrlichia ruminantium.
ORGANISM      Ehrlichia ruminantium.
REFERENCE      1 (bases 1 to 3541)
AUTHORS      Sulsona,C.R., Mahan,S.M. and Barbet,A.F.
TITLE      The map1 Gene of Cowdria ruminantium Is a Member of a Multigene
          Family Containing Both Conserved and Variable Genes
JOURNAL      Biochem. Biophys. Res. Commun. 257 (2), 300-305 (1999)
REFERENCE      2 (bases 1 to 3541)
AUTHORS      Sulsona,C.R., Mahan,S.M. and Barbet,A.F.
TITLE      Direct Submission
JOURNAL      Submitted (02-FEB-1999) College of Veterinary Medicine Department
          of Pathobiology, University of Florida, P.O. Box 110880,
          Gainesville, FL 32610, USA
FEATURES
source      Location/Qualifiers
1..3541
/organism="Ehrlichia ruminantium"
/isolate="Antigua"
/db_xref="taxon:779"
989..1837
/note="ORF2"
/codon_start=1
/transl_table=1
/product="major antigenic protein 1 like protein"
/protein_id="AAD26353.1"
/db_xref="GI:4589110"
/translation="MNYKKILVRSALISLMSFLPQSFAPVSSNNIGNENAKGFGYI
SAKNYSPHPRKFSABETPVYKDSPTKKVFLGKKGSIYKSDFTRTDISFEGQNN
FISGSGIGYIMDGPVEIAAYQKFNPKNPANETDTSYKYHGLSRAAMADKKY
VLTNNGVTFSLMFNACVDITAEGVPIPVACAGIGADLISIFDDINLKFAYOQKIG
ISYPTTPEISAFIGGYHGVGNKYNKVPVKLPVLTDPQSTASVTLDAQYFGGEL
GVRFTF"
2210..3055
/ gene="MAPI"
CDS      2210..3055
/ gene="MAPI"
/codon_start=1
/transl_table=1
/product="major antigenic protein 1"
/protein_id="AAD26354.1"
/db_xref="GI:4589111"
/translation="MNCKKIFITSTLISLVSFLPGVSFSDVIQEDSSPVGVSYSYSAKY

```

MPTASHFGKMSIKEDSRDVKVFLGKDWGVKTSSTNTIFTEKDYSEKVENNPFLOF  
AGAIGYMNQPRIEFISYETFDVKNPGNKNDAHMYCALDTSSTGGAAASTVWV  
KNEMLTISMLNACYDMLDGMFVSPYVACAGICTDLVSNTPKLSYQKLGISY  
SINPEASIFIGGHHFRVIGNEFKDITTSKIPTNTSGATPGFASAILDVCHFIEIG  
GRFVF"

BASE COUNT 1282 a 451 c 479 g 1329 t  
ORIGIN

Query Match 2.7%; Score 23; DB 1; Length 3541;  
Best Local Similarity 100.0%; Pred. No. 0.66;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 TTACCATATCAGTCTTTTGCGA 77  
|||||  
Db 1043 TTACCATATCAGTCTTTTGCGA 1065

RESULT 10  
AF125275 3551 bp DNA linear BCT 20-APR-1999  
LOCUS  
DEFINITION Cowdria ruminantium isolate Crystal Springs major antigenic protein 1 like protein and major antigenic protein 1 (MAP1) genes, complete cds.

ACCESSION AF125275 GI:4589097  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Ehrlichia ruminantium.  
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
Anaplasmataceae; Ehrlichia.

REFERENCE 1 (bases 1 to 3551)

Sulsona, C.R., Mahan, S.M. and Barbet, A.F.

The map1 Gene of Cowdria ruminantium Is a Member of a Multigene Family Containing Both Conserved and Variable Genes

Biochem. Biophys. Res. Commun. 257 (2), 300-305 (1999)

REFERENCE 2 (bases 1 to 3551)

Sulsona, C.R., Mahan, S.M. and Barbet, A.F.

Direct Submission  
Submitted (02-FEB-1999) College of Veterinary Medicine Department of Pathobiology, University of Florida, P.O. Box 110880, Gainesville, FL 32610, USA

FEATURES Location/Qualifiers

1..3551  
/organism="Ehrlichia ruminantium"  
/isolate="Crystal Springs"  
/db\_xref="taxon:779"

CDS

991..1839  
/note="ORF2"  
/codon\_start=1  
/transl\_table=11  
/product="major antigenic protein 1 like protein"  
/protein\_id="AAD26345.1"  
/db\_xref="GI:4589098"  
/translation="MNKKILVRSALISLMSFLPYQSFAPVSNNGENNAKEGFYI  
SAKNPSIPHRKFSAEETPVYKQDPTKKVFLGKKGSTIKYSDFTRTIDISFGQNN  
FISFGSGSIGYIMDGRVEIAAYQKFNPNANETDTSYKHGYSRAETMDKKY  
VLTNGVTSSLNFACYDITAEGVFPFIYACAGIGADLISIFDDINLKFAYQGGIG  
ISYPTPEISAFIGGYHGVGNKYNKVPVKLPVTLTDAPOSTASVTLDAGYFGGEL  
GVRFTF"

gene

2199..3062

/gene="MAP1"

CDS

2199..3062

/gene="MAP1"

/codon\_start=1  
/transl\_table=11  
/product="major antigenic protein 1"

/protein\_id="AAD26346.1"

/db\_xref="GI:4589099"

/translation="MNCKKIFITSLISLVSFLPGVSFSDVIOEDSNPAGSVIYSAKY  
MPTASHFGKMSIKEDSKNTQTVFLGKDWGVKTPSDSNNTNFTIFTEKDYSPRYENN  
PFLGAGAGYMNQPRIEFVSVETFDVKNPGNKNDAHMYCALDTAQNSTNGAG  
LTVSNVKNENLTNLSMLNACYDMLDGIPIVSPYVACAGICTDLVSNTPKLSYQ  
GKLGISYINSEASIFIGGHHFRVIGNEFKDITTSKIPTNTSGATPGFASAILDVCH

FGIEIGRFRVF"  
BASE COUNT 1308 a 477 c 464 g 1302 t  
ORIGIN

Query Match 2.7%; Score 23; DB 1; Length 3551;  
Best Local Similarity 100.0%; Pred. No. 0.66;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 TTACCATATCAGTCTTTTGCGA 77  
|||||  
Db 1045 TTACCATATCAGTCTTTTGCGA 1067

RESULT 11  
AF125278

LOCUS

DEFINITION

AF125278

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Ehrlichia ruminantium.

Bacteria; Proteobacteria;

REFERENCE 1 (bases 1 to 3572)

Sulsona, C.R., Mahan, S.M. and Barbet, A.F.

The map1 Gene of Cowdria ruminantium Is a Member of a Multigene Family Containing Both Conserved and Variable Genes

Biochem. Biophys. Res. Commun. 257 (2), 300-305 (1999)

REFERENCE 2 (bases 1 to 3572)

Sulsona, C.R., Mahan, S.M. and Barbet, A.F.

Direct Submission

Submitted (02-FEB-1999) College of Veterinary Medicine Department of Pathobiology, University of Florida, P.O. Box 110880, Gainesville, FL 32610, USA

FEATURES Location/Qualifiers

1..3572

/organism="Ehrlichia ruminantium"

/isolate="UmBanein"

/db\_xref="taxon:779"

989..1837

/note="ORF2"

/codon\_start=1  
/transl\_table=11  
/product="major antigenic protein 1 like protein"

/protein\_id="AAD26351.1"

/db\_xref="GI:4589107"

/translation="MNKKILVRSALISLMSFLPYQSFAPVSNNGENNAKEGFYI  
SAKNPSIPHRKFSAEETPVYKQDPTKKVFLGKKGSTIKYSDFTRTIDISFGQNN  
FISFGSGSIGYIMDGRVEIAAYQKFNPNANETDTSYKHGYSRAEAMADKKY  
VLTNGVTSSLNFACYDITAEGVFPFIYACAGIGADLISIFDDINLKFAYQGGIG  
ISYPTPEISAFIGGYHGVGNKYNKVPVKLPVTLTDAPOSTASVTLDAGYFGGEL  
GVRFTF"

2210..3082

/gene="MAP1"

2210..3082

/gene="MAP1"

/codon\_start=1  
/transl\_table=11  
/product="major antigenic protein 1"

/protein\_id="AAD26352.1"

/db\_xref="GI:4589108"

/translation="MNCKKIFITSLISLVSFLPGVSFSDVIOEDSNPAGSVIYSAKY  
MPTASHFGKMSIKEDSKNTQTVFLGKDWGVKTPSDSNNTNFTIFTEKDYSPRYENN  
PFLGAGAGYMNQPRIEFVSVETFDVKNPGNKNDAHMYCALDTGTFSGQAT  
LNSVVMKNENLTDLIALNACYDITLGMFVSPYVACAGICTDLVSNTPKLSYQ  
GKLGISYINSEASIFIGGHHFRVIGNEFKDITTSKIPTNTSGATPGFASAILD  
VCHFGIEIGRFRVF"

BASE COUNT 1303 a 472 c 476 g 1321 t

ORIGIN

Query Match 2.7%; Score 23; DB 1; Length 3572;

```

Best Local Similarity 100.0%; Pred. No. 0.66;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 TTACCATATCAGTCTTTTGCAGA 77
    |||||
Db 1043 TTACCATATCAGTCTTTTGCAGA 1065

RESULT 12
AC100255 63124 bp DNA linear HTG 22-NOV-2001
LOCUS Mus musculus clone RP23-66A22, LOW-PASS SEQUENCE SAMPLING.
DEFINITION
AC100255
AC100255.1 GI:17047621
KEYWORDS HTG; HTGS PHASE0.
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 63124)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
2 Mus musculus, clone RP23-66A22
Unpublished
2 (bases 1 to 63124)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Baschien,V., Boguslavskiy,L., Boukigalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collimore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferrelira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heatford,A., Horton,L., Hulle,W., Iliev,I., Johnson,R.,
Jones,C., Kanat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L14466
Center clone name: 66_A_22

-----
* NOTE: This record contains 79 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 696: contig of 696 bp in length

```

```

697 796: gap of 100 bp
777 1357: contig of 561 bp in length
1358 1457: gap of 100 bp
1458 2159: contig of 702 bp in length
2160 2259: gap of 100 bp
2260 2965: contig of 706 bp in length
2966 3065: gap of 100 bp
3066 3773: contig of 714 bp in length
3780 3879: gap of 100 bp
3880 4566: contig of 687 bp in length
4567 5368: contig of 702 bp in length
5369 5468: gap of 100 bp
5469 6161: contig of 693 bp in length
6162 6261: gap of 100 bp
6262 6965: contig of 704 bp in length
6966 7065: gap of 100 bp
7066 7773: contig of 708 bp in length
7774 7873: gap of 100 bp
7874 8573: contig of 706 bp in length
8580 8679: gap of 100 bp
8680 9377: contig of 698 bp in length
9378 9477: gap of 100 bp
9478 10180: contig of 703 bp in length
10181 10280: gap of 100 bp
10281 10993: contig of 713 bp in length
10994 11093: gap of 100 bp
11094 11742: contig of 649 bp in length
11743 11842: gap of 100 bp
11843 12559: contig of 717 bp in length
12560 12659: gap of 100 bp
12660 13375: contig of 716 bp in length
13376 13475: gap of 100 bp
13476 14181: contig of 706 bp in length
14182 14281: gap of 100 bp
14282 14987: contig of 706 bp in length
14988 15087: gap of 100 bp
15088 15804: contig of 717 bp in length
15805 15904: gap of 100 bp
15905 16608: contig of 704 bp in length
16609 16708: gap of 100 bp
16709 17415: contig of 707 bp in length
17416 17515: gap of 100 bp
17516 18193: contig of 677 bp in length
18193 18292: gap of 100 bp
18293 19004: contig of 712 bp in length
19005 19104: gap of 100 bp
19105 19808: contig of 704 bp in length
19809 19908: gap of 100 bp
19909 20621: contig of 713 bp in length
20622 20721: gap of 100 bp
20722 21418: contig of 697 bp in length
21419 21518: gap of 100 bp
21519 22233: contig of 715 bp in length
22234 22333: gap of 100 bp
22334 23034: contig of 701 bp in length
23035 23134: gap of 100 bp
23135 23852: contig of 718 bp in length
23853 23952: gap of 100 bp
23953 24660: contig of 708 bp in length
24661 24760: gap of 100 bp
24761 25452: contig of 692 bp in length
25453 25552: gap of 100 bp
25553 26274: contig of 722 bp in length
26275 26374: gap of 100 bp
26375 27061: contig of 687 bp in length
27062 27161: gap of 100 bp
27162 27861: contig of 700 bp in length
27862 27961: gap of 100 bp
27962 28649: contig of 688 bp in length
28650 28749: gap of 100 bp
28750 29459: contig of 710 bp in length
29460 29559: gap of 100 bp

```

\* 29560 30261: contig of 702 bp in length  
\* 30262 30361: gap of 100 bp  
\* 30362 31070: contig of 709 bp in length  
\* 31071 31170: gap of 100 bp  
\* 31171 31881: contig of 711 bp in length  
\* 31882 31981: gap of 100 bp  
\* 31982 32684: contig of 703 bp in length  
\* 32685 32784: gap of 100 bp  
\* 32785 33469: contig of 685 bp in length  
\* 33470 33569: gap of 100 bp  
\* 33570 34276: contig of 707 bp in length  
\* 34277 34376: gap of 100 bp  
\* 34377 35074: contig of 698 bp in length  
\* 35075 35174: gap of 100 bp  
\* 35175 35887: contig of 713 bp in length  
\* 35888 35987: gap of 100 bp  
\* 35988 36697: contig of 710 bp in length  
\* 36698 36797: gap of 100 bp  
\* 36798 37510: contig of 713 bp in length  
\* 37511 37610: gap of 100 bp  
\* 37611 38295: contig of 685 bp in length  
\* 38296 38395: gap of 100 bp  
\* 38396 39106: contig of 711 bp in length  
\* 39107 39206: gap of 100 bp  
\* 39207 39892: contig of 686 bp in length  
\* 39893 39992: gap of 100 bp  
\* 39993 40707: contig of 715 bp in length  
\* 40708 40807: gap of 100 bp  
\* 40808 41496: contig of 689 bp in length  
\* 41497 41596: gap of 100 bp  
\* 41597 42295: contig of 699 bp in length  
\* 42296 42395: gap of 100 bp  
\* 42396 43063: contig of 668 bp in length  
\* 43064 43163: gap of 100 bp  
\* 43164 43868: contig of 705 bp in length  
\* 43869 43968: gap of 100 bp  
\* 43969 44679: contig of 711 bp in length  
\* 44680 44779: gap of 100 bp  
\* 44779 45479: contig of 700 bp in length  
\* 45480 45579: gap of 100 bp  
\* 45580 46284: contig of 705 bp in length  
\* 46285 46384: gap of 100 bp  
\* 46385 47085: contig of 701 bp in length  
\* 47086 47185: gap of 100 bp  
\* 47186 47855: contig of 670 bp in length  
\* 47856 47955: gap of 100 bp  
\* 47956 48672: contig of 717 bp in length  
\* 48673 48772: gap of 100 bp  
\* 48773 49472: contig of 700 bp in length  
\* 49473 49572: gap of 100 bp  
\* 49573 50286: contig of 714 bp in length  
\* 50287 50386: gap of 100 bp  
\* 50387 51094: contig of 708 bp in length  
\* 51095 51194: gap of 100 bp  
\* 51195 51897: contig of 703 bp in length  
\* 51898 51997: gap of 100 bp  
\* 51998 52707: contig of 710 bp in length  
\* 52708 52807: gap of 100 bp  
\* 52808 53521: contig of 714 bp in length  
\* 53522 53621: gap of 100 bp  
\* 53622 54326: contig of 705 bp in length  
\* 54327 54426: gap of 100 bp  
\* 54427 55120: contig of 694 bp in length  
\* 55121 55220: gap of 100 bp

Query Match 2.7%; Score 23; DB 2; Length 63124;  
Best Local Similarity 100.0%; Pred. No. 0.5;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 600 TATTTTAAAGACCTCAATCTAA 622

Db 52097 TATTTTAAAGACCTCAATCTAA 52119

## RESULT 13

AL139383

LOCUS

DEFINITION

Human DNA sequence from clone RP11-87G1 on chromosome 13q12.2-13.2, complete sequence.

AL139383

VERSION

AL139383.14 GI:11139893

KEYWORDS

HTG.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1

(bases 1 to 182917)

AUTHORS

Wall, M.

TITLE

Direct Submission

JOURNAL

Submitted (11-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk requests: clonerequest@sanger.ac.uk

COMMENT

On Nov 13, 2000 this sequence version replaced GI:11137642.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence has been finished according to sequence map criteria

as follows. An attempt is made to resolve all sequencing problems,

such as compressions and repeats, but not necessarily within known

annotated human repeat sequence elements (e.g. Alu). Where the

sequence is ambiguous, there is an annotation using the 'unsure'

feature key.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

chromosome 13, constructed by the Sanger Centre Chromosome 13

Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr13

RP11-87G1 is from the library RPC1-11.1 constructed at the Roswell

Park Cancer Institute by the group of Pieter de Jong. For further

details see http://bacpac.med.buffalo.edu/

VBCOR: pBACe3.6

This sequence is the entire insert of clone RP11-87G1 The true left

end of clone RP11-37L2 is at 122531 in this sequence. The true

right end of clone RP11-141M1 is at 104969 in this sequence.

Location/Qualifiers

1. 182917

/organism="Homo sapiens"

/db xref="taxon:9606"

/chromosome="13"

/map="q12.2-13.2"

/clone="RP11-87G1"

/clone lib="RPC1-11.1"

306. 618

/note="AluX repeat: matches 1. 311 of consensus"

1312. 1416

/note="MIR repeat: matches 44. 146 of consensus"

3221. 4161

/note="L1MB3 repeat: matches 5214. 6181 of consensus"

5098. 5140

/note="MER94 repeat: matches 63. 94 of consensus"

5141. 5538

/note="THE1C repeat: matches 1. 369 of consensus"

5539. 5588

/note="MER94 repeat: matches 10. 63 of consensus"

5931. 6066

/note="L2 repeat: matches 2625. 2750 of consensus"

6185. 6331

/note="FAM repeat: matches 18. 171 of consensus"

6396. 6654

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

```
repeat_region /note="L2 repeat: matches 1346. .1650 of consensus"
8207. .8517 /note="AluSg repeat: matches 1. .310 of consensus"
8526. .8826 /note="AluJb repeat: matches 1. .301 of consensus"
9346. .9524 /note="MIR repeat: matches 6. .188 of consensus"
9577. .9662 /note="L2 repeat: matches 2641. .2731 of consensus"
9802. .10130 /note="AluSx repeat: matches 1. .300 of consensus"
10446. .10660 /note="L2 repeat: matches 2234. .2439 of consensus"
11156. .11466 /note="AluSx repeat: matches 1. .311 of consensus"
12132. .12277 /note="L2 repeat: matches 2554. .2700 of consensus"
12949. .13085 /note="FLAM C repeat: matches 1. .125 of consensus"
13160. .13480 /note="AluSx repeat: matches 1. .310 of consensus"
14663. .14842 /note="2 copies 90 mer 97% conserved"
14917. .14956 /note="20 copies 2 mer at 100% conserved"
15166. .15189 /note="L2 copies 2 mer aa 100% conserved"
15276. .15575 /note="AluSx repeat: matches 1. .309 of consensus"
16713. .17001 /note="AluJb repeat: matches 1. .295 of consensus"
17090. .17195 /note="L2 repeat: matches 2588. .2695 of consensus"
17236. .17673 /note="LTR32 repeat: matches 17. .471 of consensus"
18650. .18955 /note="AluSg repeat: matches 1. .304 of consensus"
22105. .22397 /note="AluSx repeat: matches 1. .293 of consensus"
23158. .23249 /note="MIR repeat: matches 94. .192 of consensus"
23982. .24185 /note="MER3 repeat: matches 1. .208 of consensus"
25531. .25845 /note="AluJb repeat: matches 6. .309 of consensus"
25869. .25993 /note="AluJb repeat: matches 1. .125 of consensus"
26851. .27288 /note="LTR40b repeat: matches 4. .462 of consensus"
27290. .27400 /note="HERVLA0 repeat: matches 1. .115 of consensus"
27403. .27727 /note="MLT1A2 repeat: matches 3. .374 of consensus"
27740. .27839 /note="HERVLA0 repeat: matches 114. .212 of consensus"
28030. .28777 /note="LIME1 repeat: matches 5010. .5830 of consensus"
28783. .29078 /note="LIME1 repeat: matches 5750. .6044 of consensus"
29111. .29423 /note="AluY repeat: matches 1. .306 of consensus"
29460. .29609 /note="LIMEB repeat: matches 5895. .6046 of consensus"
29644. .30157 /note="LIMEA9 repeat: matches 5673. .6279 of consensus"
30164. .30460 /note="AluY repeat: matches 1. .295 of consensus"
30473. .30570 /note="LIME2 repeat: matches 6047. .6155 of consensus"
30621. .30695 /note="LIMEC4 repeat: matches 6848. .6923 of consensus"
31016. .31473 /note="LTR16B repeat: matches 9. .464 of consensus"
```

```
repeat_region 32153. .32975 /note="LTR1 repeat: matches 5. .785 of consensus"
34472. .34774 /note="AluSg repeat: matches 1. .303 of consensus"
35593. .35867 /note="MIR repeat: matches 1. .262 of consensus"
36065. .36687 /note="MER82 repeat: matches 8. .653 of consensus"
37843. .38276 /note="LIMEB2 repeat: matches 5731. .6164 of consensus"
38406. .38904 /note="L2 repeat: matches 2247. .2744 of consensus"
38904. .39146 /note="MIR repeat: matches 6. .249 of consensus"
41357. .41500 /note="LIME1 repeat: matches 5827. .5972 of consensus"
41513. .43185 /note="LIMEC repeat: matches 1801. .3214 of consensus"
43182. .43447 /note="LIME1 repeat: matches 5128. .5395 of consensus"
43448. .43760 /note="AluSx repeat: matches 1. .312 of consensus"
43761. .44505 /note="LIME1 repeat: matches 5395. .6143 of consensus"
44786. .45091 /note="AluY repeat: matches 2. .307 of consensus"
46288. .47635 /note="LIMEB8 repeat: matches 4856. .6167 of consensus"
47669. .53793 /note="LIMEA6 repeat: matches 11. .6143 of consensus"
53796. .53865 /note="35 copies 2 mer aa 67% conserved"
53869. .53972 /note="MIR repeat: matches 20. .133 of consensus"
57285. .57351 /note="MER33 repeat: matches 2. .68 of consensus"
57372. .57671 /note="AluSg repeat: matches 1. .301 of consensus"
57674. .58355 /note="LIME1 repeat: matches 4789. .5401 of consensus"
58351. .59011 /note="LIME1 repeat: matches 5769. .6123 of consensus"
59013. .59236 /note="AluY repeat: matches 1. .290 of consensus"
59237. .59616 /note="LIME1 repeat: matches 6123. .6304 of consensus"
59617. .59914 /note="LIMEA4 repeat: matches 5477. .6256 of consensus"
59915. .60090 /note="AluY repeat: matches 53. .324 of consensus"
60107. .60873 /note="LIME1 repeat: matches 67. .191 of consensus"
60934. .61200 /note="MIR repeat: matches 67. .191 of consensus"
61349. .61469 /note="MIR repeat: matches 67. .191 of consensus"
63769. .64079 /note="MIR repeat: matches 67. .191 of consensus"
```

Query Match 2.6%; Score 22; DB 9; Length 182917;  
Best Local Similarity 100.0%; Pred.No.1.6;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 TTGATTTTCAAAATAACTTAAT 302  
|||||  
Db 165259 TTGATTTTCAAAATAACTTAAT 165280

RESULT 14  
AC124404/c AC124404 206746 bp DNA linear HTG 20-JUN-2002  
LOCUS Mus musculus chromosome UNK clone RP24-371M10, WORKING DRAFT  
DEFINITION SEQUENCE, 10 unordered pieces.  
ACCESSION AC124404

```

VERSION      AC124404.2  GI:21490656
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE       house mouse.
ORGANISM     Mus musculus

REFERENCE
AUTHORS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL      1 (bases 1 to 206746)
REFERENCE    The sequence of Mus musculus clone
AUTHORS      McPherson,J.D. and Waterston,R.H.
TITLE        Unpublished
JOURNAL      2 (bases 1 to 206746)
REFERENCE    McPherson,J.D. and Waterston,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park
AUTHORS      Parkway, St. Louis, MO 63108, USA
TITLE        3 (bases 1 to 206746)
JOURNAL      McPherson,J.D. and Waterston,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park
AUTHORS      Parkway, St. Louis, MO 63108, USA
COMMENT      On Jun 20, 2002 this sequence version replaced gi:21426523.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M.BB0371M10
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 201404 bases at least Q40
Consensus quality: 202760 bases at least Q30
Consensus quality: 203759 bases at least Q20
Insert size: 159000; agarose-fp
Quality coverage: 18.08 in Q20 bases; agarose-fp
Quality coverage: 11.50 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1342: contig of 1342 bp in length
* 1343 1442: gap of unknown length
* 1443 3039: contig of 1597 bp in length
* 3040 3139: gap of unknown length
* 3140 5230: contig of 2091 bp in length
* 5231 5330: gap of unknown length
* 5331 8327: contig of 2997 bp in length
* 8328 8427: gap of unknown length
* 8428 12632: contig of 4205 bp in length
* 12633 12733: gap of unknown length
* 12733 19830: contig of 7098 bp in length
* 19831 19931: gap of unknown length
* 19931 40389: contig of 20458 bp in length
* 40389 40489: gap of unknown length
* 40489 81665: contig of 41077 bp in length
* 81666 129045: contig of 47380 bp in length
* 129046 129145: gap of unknown length
* 129146 206746: contig of 77601 bp in length.
* Location/Qualifiers
* 1. .206746

/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-371M10"
1. .1342
/note="assembly_name:Contig13"
1443. .3039
/note="assembly_name:Contig14"
3140. .5230
/note="assembly_name:Contig15"
5331. .8327
/note="assembly_name:Contig16"
8428. .12632
/note="assembly_name:Contig17"
12733. .19830
/note="assembly_name:Contig18"
19931. .40388
/note="assembly_name:Contig19"
40489. .81565
/note="assembly_name:Contig20"
81666. .129045
/note="assembly_name:Contig21"
129146. .206746
/note="assembly_name:Contig22"

BASE COUNT 59192 a 41779 c 42010 g 62840 t 925 others
ORIGIN

Query Match 2.6% Score 22; DB 2; Length 206746;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 271 GCTCCAGGCATTGATTTTCAA 292
Db 155032 GCTCCAGGCATTGATTTTCAA 155011

RESULT 15
AC094268/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-3L7, *** SEQUENCING IN PROGRESS ***,
41 unordered pieces.
AC094268
VERSION AC094268.5 GI:21717861
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 73435)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Garrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karleson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissegh,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,

```

```

Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokoko, S., Ogih, M., Okwuonu, G.,
Oragunye, N., Oviado, R., Pace, A., Payton, B., Peary, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
Schere, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I.,
Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H., Tang, H.,
Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Thomas, S.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Wang, Q.,
Usmani, K., Vaquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
Weinstock, G., and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 73435)
Worley, K.C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 73435)
Worley, K.C.
Direct Submission
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 10, 2002 this sequence version replaced gi:20147798.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GAHN
Center clone name: CH230-3L7
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 31421 bases at least Q40
Consensus quality: 34338 bases at least Q30
Consensus quality: 36108 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 41 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1092: contig of 1092 bp in length
* 1093 1192: gap of unknown length
* 1193 2923: contig of 1731 bp in length
* 2924 3023: gap of unknown length
* 3024 4399: contig of 1376 bp in length
* 4400 4500: gap of unknown length
* 4500 5853: contig of 1354 bp in length
* 5854 5954: gap of unknown length
* 5954 7166: contig of 1213 bp in length
* 7167 7267: gap of unknown length
* 7267 8467: contig of 1201 bp in length
* 8468 8567: gap of unknown length
* 8568 9736: contig of 1169 bp in length
* 9737 9836: gap of unknown length
* 9837 11317: contig of 1481 bp in length

```

```

11318 11417: gap of unknown length
11418 12421: contig of 1004 bp in length
12422 12521: gap of unknown length
12521 12551: contig of 1232 bp in length
12552 13853: gap of unknown length
13854 14869: contig of 1016 bp in length
14870 14969: gap of unknown length
14970 16151: contig of 1182 bp in length
16152 16251: gap of unknown length
16252 17286: contig of 1035 bp in length
17287 17881: gap of unknown length
17882 18910: contig of 1424 bp in length
18911 20383: contig of 1473 bp in length
20384 21631: contig of 1148 bp in length
21632 21731: gap of unknown length
21732 22736: contig of 1005 bp in length
22737 22836: gap of unknown length
22837 24302: contig of 1466 bp in length
24303 24402: gap of unknown length
24403 25704: contig of 1302 bp in length
25705 25804: gap of unknown length
25805 27261: contig of 1457 bp in length
27262 27361: gap of unknown length
27362 28385: contig of 1024 bp in length
28386 28486: gap of unknown length
28487 30410: contig of 1925 bp in length
30411 32586: contig of 2076 bp in length
32587 32687: gap of unknown length
32688 34217: contig of 1531 bp in length
34218 34317: gap of unknown length
34318 35808: contig of 1491 bp in length
35809 36917: contig of 1009 bp in length
36918 37018: gap of unknown length
37019 38514: contig of 1497 bp in length
38515 38614: gap of unknown length
38615 40497: contig of 1882 bp in length
40498 40597: gap of unknown length
40598 42126: contig of 1530 bp in length
42127 42226: gap of unknown length
42227 43751: contig of 1525 bp in length
43752 43851: gap of unknown length
43852 45917: contig of 2066 bp in length
45918 46017: gap of unknown length
46018 48340: contig of 2323 bp in length
48341 48440: gap of unknown length
48441 51546: contig of 3106 bp in length
51547 53646: gap of unknown length
53647 53831: contig of 2185 bp in length
53832 53931: gap of unknown length
53932 55651: contig of 1720 bp in length
55652 55751: gap of unknown length
55752 57675: contig of 1924 bp in length
57676 57775: gap of unknown length
57776 59491: contig of 1716 bp in length
59492 59591: gap of unknown length
59592 62450: contig of 2859 bp in length
62451 62550: gap of unknown length
62551 65905: contig of 3355 bp in length
65906 66005: gap of unknown length
66006 70117: contig of 4112 bp in length
70118 70217: gap of unknown length
70218 73435: contig of 3218 bp in length.

```

FEATURES  
source

1. 73435  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-3L7"

BASE COUNT 20584 a 14145 c 13508 g 21179 t 4019 others  
ORIGIN



Query Match 2.5%; Score 21; DB 2; Length 73435;  
 Best Local Similarity 100.0%; Pred. No. 5.9;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 708 TATTGGTAATAATTGAGAA 728  
 DB 66923 TATTGGTAATAATTGAGAA 66903

RESULT 16  
 LOCUS AL157828 100927 bp DNA linear PRI 21-DEC-2000  
 DEFINITION Human DNA sequence from clone RP11-214120 on chromosome 9p22.3-24,  
 complete sequence.

ACCESSION AL157828  
 VERSION AL157828.14 GI:11967478

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 100927)

REFERENCE Direct Submission

AUTHORS Submitted (20-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,

TITLE CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone

JOURNAL requests: clonerequest@sanger.ac.uk

COMMENT On Dec 22, 2000 this sequence version replaced gi:11878437.

During sequence assembly, data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence has been finished according to sequence map criteria

as follows. An attempt is made to resolve all sequencing problems,

such as compressions and repeats, but not necessarily within known

annotated repeat sequence elements. Where the sequence is

ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em, ENBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping

Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr9

RP11-214120 is from the library RPCI-11.1 constructed by the group

of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone

RP11-214120 it may be shorter because we sequence overlapping

sections only once, except for a 100 base overlap.

The true right end of clone RP11-214120 is at 100927 in this

sequence. The true left end of clone RP11-187K14 is at 45712 in

this sequence. The true right end of clone RP11-318 is at 100 in

this sequence.

Location/Qualifiers

1..100927

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="9"

/map="p22.3-24"

/clone="RP11-214120"

/clone\_lib="RPCI-11.1"

3..1953

/note="L1M1 repeat: matches -1389..920 of consensus"

2067..2180

/note="MIR repeat: matches 152..261 of consensus"

2553..2586

/note="18 copies 2 mer at 97% conserved"  
 2555..2590  
 /note="9 copies 4 mer at 97% conserved"  
 3327..3362  
 /note="9 copies 4 mer at 83% conserved"  
 3838..4127  
 /note="Alusx repeat: matches 1..299 of consensus"  
 4141..4216  
 /note="38 copies 2 mer aa 75% conserved"  
 4156..4219  
 /note="16 copies 4 mer aaga 87% conserved"  
 4851..5067  
 /note="AluDo repeat: matches 65..273 of consensus"  
 5098..5159  
 /note="31 copies 2 mer aa 72% conserved"  
 6582..6613  
 /note="16 copies 2 mer aa 84% conserved"  
 7520..7642  
 /note="L1MD3 repeat: matches 7636..7977 of consensus"  
 7699..7722  
 /note="12 copies 2 mer ac 100% conserved"  
 11123..11273  
 /note="L2 repeat: matches 2323..2455 of consensus"  
 12764..12983  
 /note="MIR repeat: matches 16..229 of consensus"  
 15006..15898  
 /note="AluY repeat: matches 1..299 of consensus"  
 15922..16177  
 /note="L1M89 repeat: matches 6015..6268 of consensus"  
 16206..16437  
 /note="MIR repeat: matches 28..258 of consensus"  
 16528..16579  
 /note="26 copies 2 mer aa 76% conserved"  
 16718..16887  
 /note="MSTA repeat: matches 255..426 of consensus"  
 18253..18665  
 /note="MSTA repeat: matches 1..426 of consensus"  
 20354..20647  
 /note="Alusx repeat: matches 3..298 of consensus"  
 20883..20958  
 /note="MER58A repeat: matches 18..98 of consensus"  
 complement(21447..21958)  
 /note="match: STS: Em:G29412"  
 complement(21565..21958)  
 /note="match: STS: Em:G25434"  
 21948..21997  
 /note="25 copies 2 mer aa 72% conserved"  
 22125..22395  
 /note="LTR37B repeat: matches 211..468 of consensus"  
 22426..22620  
 /note="LTR37B repeat: matches 12..215 of consensus"  
 22980..23052  
 /note="AluY repeat: matches 234..306 of consensus"  
 23060..23277  
 /note="AluY repeat: matches 80..286 of consensus"  
 23278..23313  
 /note="9 copies 4 mer ctac 91% conserved"  
 23440..23556  
 /note="MIR repeat: matches 21..142 of consensus"  
 23563..24334  
 /note="L1MCB repeat: matches 553..1295 of consensus"  
 24335..24476  
 /note="L1M3 repeat: matches 6036..6176 of consensus"  
 24477..24865  
 /note="L1MCB repeat: matches 1426..553 of consensus"  
 24866..25273  
 /note="MSTB repeat: matches 1..426 of consensus"  
 25274..26028  
 /note="L1MCB repeat: matches 701..1427 of consensus"  
 26072..26535  
 /note="L1MCB repeat: matches 102..587 of consensus"  
 27700..27729  
 /note="MER5A repeat: matches 12..34 of consensus"

```

repeat_region 27730..28277
/note="MT1F repeat: matches 1..541 of consensus"
repeat_region 28278..28315
/note="MERSA repeat: matches 34..74 of consensus"
repeat_region 29145..30166
/note="MER63D repeat: matches 3..1061 of consensus"
repeat_region 32391..32472
/note="MIR repeat: matches 51..132 of consensus"
repeat_region 33312..33619
/note="L2 repeat: matches 2439..2750 of consensus"
repeat_region 35680..35968
/note="AluSg repeat: matches 1..290 of consensus"
repeat_region 36231..36533
/note="AluY repeat: matches 2..293 of consensus"
repeat_region 36809..36949
/note="MIR repeat: matches 103..247 of consensus"
misc_feature complement(37814..38132)
/note="match: STS: Em:G23397"
repeat_region 39045..39243
/note="MIR repeat: matches 37..257 of consensus"
repeat_region 40498..40772
/note="AluY repeat: matches 35..307 of consensus"
repeat_region 40804..41094
/note="AluJb repeat: matches 14..307 of consensus"
repeat_region 41123..41250
/note="FIAM_A repeat: matches 1..128 of consensus"
repeat_region 41295..42538
/note="L2 repeat: matches 1381..2739 of consensus"
repeat_region 42565..42626
/note="MIR repeat: matches 206..262 of consensus"
repeat_region 42577..42684
/note="L2 repeat: matches 2587..2693 of consensus"
repeat_region 42694..42973
/note="MT1D repeat: matches 3..302 of consensus"
repeat_region 44082..44248
/note="MIR repeat: matches 85..261 of consensus"
repeat_region 46047..47118
/note="HSMAR1 repeat: matches 1..1287 of consensus"
repeat_region 49229..49523
/note="AluSc repeat: matches 1..297 of consensus"
repeat_region 49791..49930
/note="Alu repeat: matches 2..41 of consensus"
repeat_region 49833..50007
/note="MER47A repeat: matches 1..186 of consensus"
repeat_region 52054..52332
/note="AluX repeat: matches 1..279 of consensus"
repeat_region 52682..52875
/note="LiMD3 repeat: matches 7430..7607 of consensus"
repeat_region 53175..53985
/note="LiME repeat: matches 445..1287 of consensus"
repeat_region 54819..56031
/note="LiMEC repeat: matches 2284..3257 of consensus"
repeat_region 56032..56392
/note="MT1A1 repeat: matches 1..363 of consensus"
repeat_region 56393..56517
/note="LiMEC repeat: matches 2159..2284 of consensus"
repeat_region 56592..56913
/note="AluJb repeat: matches 1..312 of consensus"
repeat_region 57151..58172
/note="LiM4C repeat: matches 742..1750 of consensus"
repeat_region 58574..58616
/note="MT1A1 repeat: matches 3..47 of consensus"
repeat_region 58617..58748
/note="FIAM_C repeat: matches 1..133 of consensus"
repeat_region 58749..59094
/note="MT1A1 repeat: matches 47..365 of consensus"

Query Match 2.5%; Score 21; DB 9; Length 100927;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 419 AATACTATAACATTTCAT 439
|||||

```

```

Db 87570 AATACTATAACATTTCAT 87590

RESULT 17
AL732388 108867 bp DNA linear ROD 22-AUG-2002
LOCUS Mouse DNA sequence from clone RP23-358M10 on chromosome 11,
DEFINITION complete sequence.
ACCESSION AL732388
VERSION AL732388.9 GI:22474415
KEYWORDS HTG.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Tromans,A.
Direct Submission
Submitted (22-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 23, 2002 this sequence version replaced gi:21690018.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep/
from the RPI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6.
Location/Qualifiers
1..108867
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-358M10"
/clone_lib="RPI-23"
BASE COUNT 31546 a 22357 c 21416 g 33548 t
ORIGIN

Query Match 2.5%; Score 21; DB 10; Length 108867;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 AACAAATTCCTCACTAAAAA 212
|||||
Db 45962 AACAAATTCCTCACTAAAAA 45982

RESULT 18
AP004725 146677 bp DNA linear HTG 21-MAR-2002
LOCUS AP004725
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 6 clone P0412C04,

```

\*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.

```

ACCESSION      AP004725
VERSION        AP004725.1  GI:18539235
KEYWORDS       HTG; HTGS_PHASE2
SOURCE         Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
               clone:P0412C04.
ORGANISM       Oryza sativa (japonica cultivar-group)
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
               Ehrhartoideae; Oryzeae; Oryza.
REFERENCE      1
AUTHORS        Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE          Oryza sativa nipponbare(GR3) genomic DNA, chromosome 6, PAC
               clone:P0412C04
JOURNAL        Published Only in Database (2002)
REFERENCE      2 (bases 1 to 146677)
AUTHORS        Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE          Direct Submission
JOURNAL        Submitted (08-FEB-2002) Takuji Sasaki, National Institute of
               Agrobiological Sciences, Rice Genome Research Program; Kannondai
               2-1-2, Tsukuba, Ibaraki 305-8602, Japan
               (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
               Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT        NOTE: It currently consists of 1 contigs. Gaps between the contigs
               are represented as runs of N. The order of the pieces is believed
               to be correct as given, however the sizes of the gaps between them
               are based on estimates that have provided by the submitter. This
               sequence will be replaced by the finished sequence as soon as it is
               available and the accession number will be preserved.
               * NOTE: This is a 'working draft' sequence.
               * This sequence will be replaced
               * by the finished sequence as soon as it is available and
               * the accession number will be preserved.
               Location/Qualifiers
               1. 146677
                  /organism="Oryza sativa (japonica cultivar-group)"
                  /cultivar="Nipponbare"
                  /db_xref="taxon:39947"
                  /chromosome="6"
                  /clones="P0412C04"
BASE COUNT     43282 a 29986 c 29413 g 43946 t 50 others
ORIGIN
Query Match    2.5%; Score 21; DB 2; Length 146677;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 764 CTCCTCAACACATCTGCTT 784
      |||||
Db 123318 CTCCTCAACACATCTGCTT 123338

RESULT 19
LOCUS      AC020680
DEFINITION Homo sapiens clone RP11-1416, WORKING DRAFT SEQUENCE, 43 unordered
            pieces.
ACCESSION  AC020680
VERSION    AC020680.6  GI:21166332
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 148919)
REFERENCE  Birren, B., Linton, L., Nussbaum, C. and Lander, E.
AUTHORS    Unpublished
TITLE      2 (bases 1 to 148919)
JOURNAL    Submitted (08-FEB-2002) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT    NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            Location/Qualifiers
            1. 146677
               /organism="Oryza sativa (japonica cultivar-group)"
               /cultivar="Nipponbare"
               /db_xref="taxon:39947"
               /chromosome="6"
               /clones="P0412C04"
BASE COUNT     43282 a 29986 c 29413 g 43946 t 50 others
ORIGIN
Query Match    2.5%; Score 21; DB 2; Length 146677;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 764 CTCCTCAACACATCTGCTT 784
      |||||
Db 123318 CTCCTCAACACATCTGCTT 123338

RESULT 19
LOCUS      AC020680
DEFINITION Homo sapiens clone RP11-1416, WORKING DRAFT SEQUENCE, 43 unordered
            pieces.
ACCESSION  AC020680
VERSION    AC020680.6  GI:21166332
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 148919)
REFERENCE  Birren, B., Linton, L., Nussbaum, C. and Lander, E.
AUTHORS    Unpublished
TITLE      2 (bases 1 to 148919)
JOURNAL    Submitted (08-FEB-2002) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT    NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            Location/Qualifiers
            1. 146677
               /organism="Oryza sativa (japonica cultivar-group)"
               /cultivar="Nipponbare"
               /db_xref="taxon:39947"
               /chromosome="6"
               /clones="P0412C04"
BASE COUNT     43282 a 29986 c 29413 g 43946 t 50 others
ORIGIN

```

DeArellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., Oliver, T.M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (08-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 148919)

REFERENCE  
AUTHORS  
Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicoll, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trifilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (24-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 24, 2002 this sequence version replaced gi:14209765.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L3407  
Center clone name: 14\_I\_6  
----- Summary Statistics  
Sequencing vector: M13; M77815; 12% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 123100 bases at least Q40  
Consensus quality: 136061 bases at least Q30  
Consensus quality: 141228 bases at least Q20  
Insert size: 188000; agarose-fp  
Insert size: 144719; sum-of-contigs  
Quality coverage: 10.9 in Q20 bases; agarose-fp  
Quality coverage: 14.1 in Q20 b.

\* NOTE: This is a 'working draft' sequence. It currently consists of 43 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will

```
* be preserved
* 1 10015: contig of 10015 bp in length
* 10016 10115: gap of 100 bp
* 10116 11346: contig of 1231 bp in length
* 11347 11446: gap of 100 bp
* 11447 12156: contig of 710 bp in length
* 12157 12256: gap of 100 bp
* 12257 13516: contig of 1260 bp in length
* 13517 13616: gap of 100 bp
* 13617 14754: contig of 1138 bp in length
* 14755 14854: gap of 100 bp
* 14855 16253: contig of 1399 bp in length
* 16254 16353: gap of 100 bp
* 16354 17689: contig of 1336 bp in length
* 17690 17789: gap of 100 bp
* 17790 19034: contig of 1245 bp in length
* 19035 19134: gap of 100 bp
* 19135 20728: contig of 1594 bp in length
* 20729 20828: gap of 100 bp
* 20829 23037: contig of 2209 bp in length
* 23038 23137: gap of 100 bp
* 23138 25388: contig of 2251 bp in length
* 25389 25488: gap of 100 bp
* 25489 27103: contig of 1615 bp in length
* 27104 27203: gap of 100 bp
* 27204 29369: contig of 2166 bp in length
* 29370 29469: gap of 100 bp
* 29470 31731: contig of 2262 bp in length
* 31732 31831: gap of 100 bp
* 31832 34520: contig of 2699 bp in length
* 34521 34620: gap of 100 bp
* 34621 36978: contig of 2358 bp in length
* 36979 37078: gap of 100 bp
* 37079 38633: contig of 1555 bp in length
* 38634 38733: gap of 100 bp
* 38734 40786: contig of 2053 bp in length
* 40787 40886: gap of 100 bp
* 40887 42903: contig of 2017 bp in length
* 42904 43003: gap of 100 bp
* 43004 44271: contig of 1268 bp in length
* 44272 44371: gap of 100 bp
* 44372 46615: contig of 2244 bp in length
* 46616 46715: gap of 100 bp
* 46716 49350: contig of 2635 bp in length
* 49351 49450: gap of 100 bp
* 49451 52535: contig of 3085 bp in length
* 52536 52635: gap of 100 bp
* 52636 56059: contig of 3424 bp in length
* 56060 56159: gap of 100 bp
* 56160 59005: contig of 2846 bp in length
* 59006 59105: gap of 100 bp
* 59106 62614: contig of 3509 bp in length
* 62615 62714: gap of 100 bp
* 62715 65316: contig of 2602 bp in length
* 65317 65416: gap of 100 bp
* 65417 76352: contig of 10936 bp in length
* 76353 76452: gap of 100 bp
* 76453 78682: contig of 2230 bp in length
* 78683 78782: gap of 100 bp
* 78783 82382: contig of 3600 bp in length
* 82383 82482: gap of 100 bp
* 82483 86313: contig of 3831 bp in length
* 86314 86413: gap of 100 bp
* 86414 90448: contig of 4035 bp in length
* 90449 90548: gap of 100 bp
* 90549 95980: contig of 5432 bp in length
* 95981 96080: gap of 100 bp
* 96081 100554: contig of 4474 bp in length
* 100555 10654: gap of 100 bp
* 10655 102594: contig of 1940 bp in length
* 102595 102694: gap of 100 bp
* 102695 107496: contig of 4802 bp in length
* 107497 107596: gap of 100 bp
```

```
* 107597 114405: contig of 6809 bp in length
* 114406 114505: gap of 100 bp
* 114506 120522: contig of 6017 bp in length
* 120523 120622: gap of 100 bp
* 120623 127387: contig of 6765 bp in length
* 127388 127487: gap of 100 bp
* 127488 132863: contig of 5376 bp in length
* 132864 132963: gap of 100 bp
* 132964 135445: contig of 2482 bp in length
* 135446 135545: gap of 100 bp
* 135546 146349: contig of 10804 bp in length
* 146350 146449: gap of 100 bp
* 146450 148919: contig of 2470 bp in length.

FEATURES
    source
        1..148919
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="RP11-1416"
            /clone_lib="RPCI-11 Human Male BAC"
        1..10015
            /note="assembly_fragment"
            /clone_end:SP6
            /vector_side:left
            10116..11346
                /note="assembly_fragment"
            11447..12156
                /note="assembly_fragment"
            12257..13516
                /note="assembly_fragment"
            13617..14754
                /note="assembly_fragment"
            14855..16253
                /note="assembly_fragment"
            16354..17689
                /note="assembly_fragment"

Query Match          2.5%; Score 21; DB 2; Length 148919;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 768 TCAAACCATCTCTGCTTCAGT 788
Db 79553 TCAAACCATCTCTGCTTCAGT 79573

RESULT 20
AC025197/c
DEFINITION Homo sapiens chromosome 11 clone RP11-209E9 map 11, WORKING DRAFT
SEQUENCE, 14 unordered pieces.
ACCESSION AC025197
VERSION AC025197.2 GI:10198530
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 151865)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abramson,H., Allen,N.,
Boguslavsky,I., Bouckhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
```

McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
 Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
 Naylor,T., Naylor,J., Nihova,T., O'Connor,T., O'Donnell,P.,  
 O'Neill,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

# TITLE

JOURNAL  
 Submitted (07-MAR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

## REFERENCE

AUTHORS  
 3 (bases 1 to 151865)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Barna,N., Bastien,V., Beka,F., Boguslavsky,L.,  
 Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,  
 Chepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
 Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,  
 FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,  
 Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,  
 Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,  
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G.,  
 MacDonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,  
 McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,  
 Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,  
 O'Donnell,P., O'Neill,D., Olivar,T.M., Oliver,J., Peterson,K.,  
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,  
 Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,  
 Sognes,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,  
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,  
 Zimmer,A. and Zody,M.

# TITLE

JOURNAL  
 Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Sep 20, 2000 this sequence version replaced gi:1188875.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 ----- Genome Center -----  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)  
 ----- Project Information -----  
 Center project name: L7490  
 Center clone name: 209\_E\_9  
 ----- Summary Statistics -----  
 Sequencing vector: M13; M7815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 14689 bases at least Q40  
 Consensus quality: 149124 bases at least Q30  
 Consensus quality: 149969 bases at least Q20  
 Insert size: 151000; agarose-fp  
 Quality coverage: 5.9 in Q20 bases; agarose-fp  
 Quality coverage: 5.9 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 14 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 9341: contig of 9341 bp in length  
 \* 9342 9441: gap of 100 bp  
 \* 9442 11373: contig of 1932 bp in length  
 \* 11374 11473: gap of 100 bp

\* 11474 47057: contig of 35584 bp in length  
 \* 47058 47157: gap of 100 bp  
 \* 47158 49227: contig of 2070 bp in length  
 \* 49228 49327: gap of 100 bp  
 \* 49328 54639: contig of 5312 bp in length  
 \* 54640 54739: gap of 100 bp  
 \* 54740 59242: contig of 4503 bp in length  
 \* 59243 59342: gap of 100 bp  
 \* 59343 65284: contig of 5942 bp in length  
 \* 65285 65384: gap of 100 bp  
 \* 65385 70923: contig of 5539 bp in length  
 \* 70924 71023: gap of 100 bp  
 \* 71024 78085: contig of 7062 bp in length  
 \* 78086 78185: gap of 100 bp  
 \* 78186 87419: contig of 9234 bp in length  
 \* 87420 87519: gap of 100 bp  
 \* 87520 97716: contig of 10197 bp in length  
 \* 97717 97816: gap of 100 bp  
 \* 97817 120620: contig of 22804 bp in length  
 \* 120621 120720: gap of 100 bp  
 \* 120721 150210: contig of 29490 bp in length  
 \* 150211 150310: gap of 100 bp  
 \* 150311 151865: contig of 1555 bp in length.

# FEATURES

source  
 1. 151865  
 /organism="Homo sapiens"  
 /db xref="taxon:9606"  
 /chromosome="11"  
 /map="11"  
 /clone="RP11-20959"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 misc\_feature  
 1. 9341  
 /note="assembly\_fragment"  
 /clone\_end:SP6  
 /vector\_side:left  
 misc\_feature  
 9442. 11373  
 /note="assembly\_fragment"  
 misc\_feature  
 11474. 47057  
 /note="assembly\_fragment"  
 misc\_feature  
 47158. 49227  
 /note="assembly\_fragment"  
 misc\_feature  
 49328. 54639  
 /note="assembly\_fragment"  
 misc\_feature  
 54740. 59242  
 /note="assembly\_fragment"  
 misc\_feature  
 59343. 65284  
 /note="assembly\_fragment"  
 misc\_feature  
 65385. 70923  
 /note="assembly\_fragment"  
 misc\_feature  
 71024. 78085  
 /note="assembly\_fragment"  
 misc\_feature  
 78186. 87419  
 /note="assembly\_fragment"  
 misc\_feature  
 87520. 97716  
 /note="assembly\_fragment"  
 misc\_feature  
 97817. 120620  
 /note="assembly\_fragment"  
 misc\_feature  
 120721. 150210  
 /note="assembly\_fragment"  
 misc\_feature  
 150311. 151865  
 /note="assembly\_fragment"  
 /clone\_end:T7  
 /vector\_side:right  
 BASE COUNT 45019 a 30571 c 30486 g 44487 t 1302 others  
 ORIGIN

Query Match 2.0%; Score 21; DB 2; Length 151865;  
 Best Local Similarity 100.0%; Pred.No. 5.5;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 767 CTCGAACCATCTGCTTCAG 787

Db 146985 CTCGAACCATCTGCTTCAG 146965

```

RESULT 21
AC112846
LOCUS
DEFINITION
AC112846 169172 bp DNA linear HTG 17-JUL-2002
***, 53 unordered pieces.
AC112846
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 169172)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooks,S.L., Amarutunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,S., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabis,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Honsi,P., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsone,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,B., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapue,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peary,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,K.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wlaczky,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
JOURNAL
TITLE
REFERENCE
2 (bases 1 to 169172)
Unpublished
Worley,K.C.
JOURNAL
TITLE
Direct Submission
Submitted (25-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
JOURNAL
TITLE
3 (bases 1 to 169172)
Direct Submission
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
JOURNAL
TITLE
On Jul 13, 2002 this sequence version replaced gi:18874820.
Center: Baylor College of Medicine

```

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GRNU

Center clone name: CH230-130E10

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 133907 bases at least Q40

Consensus quality: 137668 bases at least Q30

Consensus quality: 141028 bases at least Q20

-----

\* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently  
 consists of 53 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 1030: contig of 1030 bp in length  
 1031 1130: gap of unknown length  
 1131 2131: contig of 1001 bp in length  
 2132 2231: gap of unknown length  
 2232 3243: contig of 1012 bp in length  
 3244 3343: gap of unknown length  
 3344 4467: contig of 1124 bp in length  
 4468 4567: gap of unknown length  
 4568 6230: contig of 1663 bp in length  
 6231 7393: contig of 1063 bp in length  
 7394 7493: gap of unknown length  
 7494 8789: contig of 1296 bp in length  
 8790 8889: gap of unknown length  
 8890 10249: contig of 1360 bp in length  
 10250 10349: gap of unknown length  
 10350 11886: contig of 1537 bp in length  
 11887 11886: gap of unknown length  
 11887 14051: contig of 2065 bp in length  
 14052 14151: gap of unknown length  
 14152 16180: contig of 2029 bp in length  
 16181 16280: gap of unknown length  
 16281 17833: contig of 1553 bp in length  
 17834 17933: gap of unknown length  
 17934 18960: contig of 1027 bp in length  
 18961 19060: gap of unknown length  
 19061 20895: contig of 1835 bp in length  
 20896 20995: gap of unknown length  
 20996 23245: contig of 2250 bp in length  
 23246 23245: gap of unknown length  
 23246 24731: contig of 1386 bp in length  
 24732 24831: gap of unknown length  
 24832 26846: contig of 2015 bp in length  
 26847 28326: contig of 1380 bp in length  
 28327 28426: gap of unknown length  
 28427 31145: contig of 2719 bp in length  
 31146 31245: gap of unknown length  
 31246 33462: contig of 2117 bp in length  
 33463 35233: contig of 1771 bp in length  
 35234 35333: gap of unknown length  
 35334 36993: contig of 1660 bp in length  
 36994 37093: gap of unknown length  
 37094 38947: contig of 1754 bp in length  
 38948 42194: contig of 3247 bp in length  
 42195 42294: gap of unknown length  
 42295 45101: contig of 2807 bp in length

\* 45102 45201: gap of unknown length  
 \* 45202 47537: contig of 2336 bp in length  
 \* 47538 47637: gap of unknown length  
 \* 47638 50377: contig of 2740 bp in length  
 \* 50378 50477: gap of unknown length  
 \* 50478 53259: contig of 2782 bp in length  
 \* 53260 53359: gap of unknown length  
 \* 53360 53047: contig of 1688 bp in length  
 \* 55048 55147: gap of unknown length  
 \* 55148 58332: contig of 3185 bp in length  
 \* 58333 58432: gap of unknown length  
 \* 58433 61899: contig of 3467 bp in length  
 \* 61900 61999: gap of unknown length  
 \* 62000 66059: contig of 4060 bp in length  
 \* 66060 66159: gap of unknown length  
 \* 66160 68490: contig of 2331 bp in length  
 \* 68491 68590: gap of unknown length  
 \* 68591 71683: contig of 3093 bp in length  
 \* 71684 71783: gap of unknown length  
 \* 71784 76624: contig of 4841 bp in length  
 \* 76625 76724: gap of unknown length  
 \* 76725 81125: contig of 4401 bp in length  
 \* 81126 81225: gap of unknown length  
 \* 81226 84264: contig of 3039 bp in length  
 \* 84265 84364: gap of unknown length  
 \* 84365 88212: contig of 3848 bp in length  
 \* 88213 88312: gap of unknown length  
 \* 88313 92555: contig of 4243 bp in length  
 \* 92556 92655: gap of unknown length  
 \* 92656 97889: contig of 5334 bp in length  
 \* 97890 98089: gap of unknown length  
 \* 98090 99966: contig of 1877 bp in length  
 \* 99967 100066: gap of unknown length  
 \* 100067 102119: contig of 2053 bp in length  
 \* 102120 102219: gap of unknown length  
 \* 102220 105229: contig of 3010 bp in length  
 \* 105230 105329: gap of unknown length  
 \* 105330 111586: contig of 6257 bp in length  
 \* 111587 111687: gap of unknown length  
 \* 111688 115834: contig of 4148 bp in length  
 \* 115835 115934: gap of unknown length  
 \* 115935 121594: contig of 5660 bp in length  
 \* 121595 121694: gap of unknown length  
 \* 121695 125201: contig of 3507 bp in length  
 \* 125202 125301: gap of unknown length  
 \* 125302 131024: contig of 5723 bp in length  
 \* 131025 131124: gap of unknown length  
 \* 131125 137260: contig of 6136 bp in length  
 \* 137261 137360: gap of unknown length  
 \* 137361 142439: contig of 5079 bp in length  
 \* 142440 142539: gap of unknown length  
 \* 142540 149482: contig of 6943 bp in length  
 \* 149483 149582: gap of unknown length  
 \* 149583 159775: contig of 10193 bp in length  
 \* 159776 159875: gap of unknown length  
 \* 159876 169172: contig of 9297 bp in length.  
 FEATURES Location/Qualifiers  
 source 1. 169172

Query Match 2.5%; Score 21; DB 2; Length 169172;  
 Best Local Similarity 100.0%; Pred. No. 5.5;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 708 TATTGGTAATAAATTGAGAA 728  
 |||||  
 Db 53592 TATTGGTAATAAATTGAGAA 53612

RESULT 22  
 AC113984 AC113984 169969 bp DNA linear HTG 06-JUN-2002  
 LOCUS Mus musculus clone RP24-97D3, WORKING DRAFT SEQUENCE, 20 ordered  
 DEFINITION pieces.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS

TITLE  
 JOURNAL  
 COMMENT

AC113984  
 AC113984.2 GI:21327615  
 HTG; HTGS PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
 house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 169969)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Mus musculus, clone RP24-97D3  
 Unpublished  
 2 (bases 1 to 169969)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouktgatier,B.,  
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Charaz,B.,  
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
 Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gaidyna,S.,  
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,  
 Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,  
 Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,  
 McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,I.,  
 Mieng,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,  
 Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
 Retta,R., Rieback,M., Riley,R., Riese,C., Rogov,P., Roman,J.,  
 Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,  
 Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (06-MAR-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 169969)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,  
 Bouktgatier,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
 Charaz,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
 Cooke,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,  
 Galagan,J., Gaidyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,  
 Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,  
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,  
 Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,  
 Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,  
 Meneus,L., Mihova,I., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
 Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Riese,C.,  
 Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,  
 Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,  
 Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,  
 Theodore,J., Topham,K., Travers,M., Travis,N., Triglio,J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jun 6, 2002 this sequence version replaced gi19172889.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information

Center project name: L24343  
 Center clone name: 97\_D\_3  
 ----- Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 162083 bases at least Q40  
 Consensus quality: 165823 bases at least Q30  
 Consensus quality: 167208 bases at least Q20  
 Insert size: 168000; agarose-ep  
 Quality coverage: 5.6 in Q20 bases; agarose-ep  
 Quality coverage: 5.6 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 20 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submittor.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

1 654: contig of 654 bp in length  
 655 754: gap of 100 bp  
 755 2003: contig of 1249 bp in length  
 2004 2103: gap of 100 bp  
 2104 3591: contig of 1488 bp in length  
 3592 3691: gap of 100 bp  
 3692 4879: contig of 1188 bp in length  
 4880 4979: gap of 100 bp  
 4980 6802: contig of 1823 bp in length  
 6803 6902: gap of 100 bp  
 6903 7890: contig of 988 bp in length  
 7891 7990: gap of 100 bp  
 7991 10607: contig of 2617 bp in length  
 10608 10707: gap of 100 bp  
 10708 12547: contig of 1840 bp in length  
 12548 12647: gap of 100 bp  
 12648 15326: contig of 2679 bp in length  
 15327 15426: gap of 100 bp  
 15427 18735: contig of 3309 bp in length  
 18736 18835: gap of 100 bp  
 18836 23759: contig of 4924 bp in length  
 23760 23859: gap of 100 bp  
 23860 27895: contig of 4036 bp in length  
 27896 27995: gap of 100 bp  
 34882 34981: contig of 6886 bp in length  
 34982 44136: contig of 9155 bp in length  
 44137 44236: gap of 100 bp  
 44237 52092: contig of 7856 bp in length  
 52093 52192: gap of 100 bp  
 52193 60084: contig of 7892 bp in length  
 60085 60184: gap of 100 bp  
 60185 76086: contig of 15902 bp in length  
 76087 76186: gap of 100 bp  
 76187 99362: contig of 23176 bp in length  
 99363 99462: gap of 100 bp  
 99463 123738: contig of 24276 bp in length  
 123739 123838: gap of 100 bp  
 123839 169969: contig of 46131 bp in length.

## FEATURES

source  
 1..169969  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="RP24-97D3"  
 /clone.lib="RPCI-24 Male Mouse BAC"  
 1..654  
 /note="assembly\_fragment"  
 755..2003  
 /note="assembly\_fragment"  
 2104..3591  
 /note="assembly\_fragment"

misc\_feature  
 3692..4879  
 /note="assembly\_fragment"  
 misc\_feature  
 4980..6802  
 /note="assembly\_fragment"  
 misc\_feature  
 6903..7890  
 /note="assembly\_fragment"  
 misc\_feature  
 7991..10607  
 /note="assembly\_fragment"  
 misc\_feature  
 10708..12547  
 /note="assembly\_fragment"  
 misc\_feature  
 12648..15326  
 /note="assembly\_fragment"  
 misc\_feature  
 15427..18735  
 /note="assembly\_fragment"  
 misc\_feature  
 18836..23759  
 /note="assembly\_fragment"  
 misc\_feature  
 23860..27895  
 /note="assembly\_fragment"  
 misc\_feature  
 27996..34881  
 /note="assembly\_fragment"  
 misc\_feature  
 34982..44136  
 /note="assembly\_fragment"  
 misc\_feature  
 44237..52092  
 /note="assembly\_fragment"  
 misc\_feature  
 52193..60084  
 /note="assembly\_fragment"  
 misc\_feature  
 60185..76086  
 /note="assembly\_fragment"  
 misc\_feature  
 76187..99362  
 /note="assembly\_fragment"  
 misc\_feature  
 99463..123738  
 /note="assembly\_fragment"  
 misc\_feature  
 123839..169969  
 /note="assembly\_fragment"  
 BASE COUNT 46367 a 38608 c 37808 g 45279 t 1907 others  
 ORIGIN  
 Query Match 2.5%; Score 21; DB 2; Length 169969;  
 Best Local Similarity 100.0%; Pred. No. 5.5;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 564 ATATGTCATGTCAGGTATAGG 584  
 Db 89142 ATATGTCATGTCAGGTATAGG 89162  
 RESULT 23  
 AC115013  
 LOCUS AC115013 180931 bp DNA linear HTG 17-JUN-2002  
 DEFINITION Mus musculus clone RP24-193P19, WORKING DRAFT SEQUENCE, 20 ordered  
 pieces.  
 ACCESSION AC115013  
 VERSION AC115013.2 GI:21431344  
 KEYWORDS HTG; HTGS PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 180931)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 TITLE Mus musculus, clone RP24-193P19  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 180931)  
 AUTHORS Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,  
 Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
 Ginde,S., Gerred,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kellis,C., LaRocque,K., Lamazares,R.,



Landers, T., Lehoczyk, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Milenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (14-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 180931)

Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhaltier, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczyk, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Milenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (17-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 17, 2002 this sequence version replaced GI:19424509.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center -----  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information -----  
Center project name: L24608  
Center clone name: 193\_P19  
----- Summary Statistics -----  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 172404 bases at least Q40  
Consensus quality: 176966 bases at least Q30  
Consensus quality: 178461 bases at least Q20  
Insert size: 182000; agarose-fp  
Insert size: 179031; sum-of-contigs  
Quality coverage: 5.6 in Q20 bases; agarose-fp  
Quality coverage: 5.7 in Q20 bases; sum-of-contigs  
----- NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter. -----  
\* This sequence will be replaced

\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

1 1632: contig of 1632 bp in length  
1633 1732: gap of 100 bp  
1733 3488: contig of 1756 bp in length  
3489 3588: gap of 100 bp  
3589 5648: contig of 2060 bp in length  
5649 5748: gap of 100 bp  
5749 7630: contig of 1882 bp in length  
7631 7730: gap of 100 bp  
7731 9842: contig of 2112 bp in length  
9843 9942: gap of 100 bp  
9943 12290: contig of 2348 bp in length  
12291 12390: gap of 100 bp  
12391 14428: contig of 2038 bp in length  
14429 14528: gap of 100 bp  
14529 17176: contig of 2648 bp in length  
17177 17276: gap of 100 bp  
17277 19834: contig of 2558 bp in length  
19835 19934: gap of 100 bp  
19935 23350: contig of 3416 bp in length  
23351 23450: gap of 100 bp  
23451 27068: contig of 3618 bp in length  
27069 27168: gap of 100 bp  
27169 31839: contig of 4671 bp in length  
31840 31939: gap of 100 bp  
31940 36047: contig of 4108 bp in length  
36048 36147: gap of 100 bp  
36148 43199: contig of 7052 bp in length  
43200 43299: gap of 100 bp  
43300 50770: contig of 7471 bp in length  
50771 50870: gap of 100 bp  
50871 64683: contig of 13813 bp in length  
64684 64783: gap of 100 bp  
64784 84745: contig of 19963 bp in length  
84747 84846: gap of 100 bp  
84847 108626: contig of 23780 bp in length  
108627 108726: gap of 100 bp  
108727 138858: contig of 30132 bp in length  
138859 138958: gap of 100 bp  
138959 180931: contig of 41973 bp in length.

FEATURES  
source  
1. 180931  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP24-193p19"  
/clone\_lib="RPCI-24 Male Mouse BAC"  
misc\_feature  
1. 1832  
/notes="assembly\_fragment"  
1733. 3488  
/notes="assembly\_fragment"  
misc\_feature  
3589. 5648  
/notes="assembly\_fragment"  
misc\_feature  
5749. 7630  
/notes="assembly\_fragment"  
7731. 9842  
/notes="assembly\_fragment"  
misc\_feature  
9943. 12290  
/notes="assembly\_fragment"  
misc\_feature  
12391. 14428  
/notes="assembly\_fragment"  
misc\_feature  
14529. 17176  
/notes="assembly\_fragment"  
misc\_feature  
17277. 19834  
/notes="assembly\_fragment"  
misc\_feature  
19935. 23350  
/notes="assembly\_fragment"  
misc\_feature  
23451. 27068  
/notes="assembly\_fragment"  
misc\_feature  
27169. 31839  
/notes="assembly\_fragment"  
misc\_feature  
31940. 36047  
/notes="assembly\_fragment"

```

misc_feature      36148..43199
                   /note="assembly_fragment"
misc_feature      43300..50770
                   /note="assembly_fragment"
misc_feature      50871..64683
                   /note="assembly_fragment"
misc_feature      64784..84746
                   /note="assembly_fragment"
misc_feature      84847..108626
                   /note="assembly_fragment"
misc_feature      108727..138858
                   /note="assembly_fragment"
misc_feature      138959..180931
                   /note="assembly_fragment"
BASE COUNT      53241 a 35193 c 35694 g 54901 t 1902 others
ORIGIN
Query Match      2.5%; Score 21; DB 2; Length 180931;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      414 TGGTGAATCACTATAAACATTT 434
        |||||
Db      170651 TGGTGAATCACTATAAACATTT 170671

RESULT 24
AC018682/c      AC018682      181405 bp      DNA      linear      PRI 09-MAY-2001
LOCUS      Homo sapiens BAC clone RP11-417F21 from 2, complete sequence.
DEFINITION
ACCESSION      AC018682
VERSION      AC018682.4 GI:11038582
KEYWORDS      HTG.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS      Kyung,K., Drone,K., Hawkins,M. and Myers,M.
TITLE      The sequence of Homo sapiens BAC clone RP11-417F21
JOURNAL      Unpublished
REFERENCE
AUTHORS      Sulston,J.E. and Waterston,R.
TITLE      Toward a complete human genome sequence
JOURNAL      Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE      99063792
PUBMED      9847074
REFERENCE
AUTHORS      2 (bases 1 to 181405)
TITLE      Kyung,K., Drone,K., Hawkins,M. and Myers,M.
JOURNAL      The sequence of Homo sapiens BAC clone RP11-417F21
REFERENCE
AUTHORS      3 (bases 1 to 181405)
TITLE      Waterston,R.H.
JOURNAL      Direct Submission
SUBMITTED      Submitted (16-DEC-1999) Genome Sequencing Center, Washington
UNIVERSITY      University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO      63108, USA
REFERENCE
AUTHORS      4 (bases 1 to 181405)
TITLE      Waterston,R.H.
JOURNAL      Direct Submission
SUBMITTED      Submitted (30-OCT-2000) Genome Sequencing Center, Washington
UNIVERSITY      University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO      63108, USA
REFERENCE
AUTHORS      5 (bases 1 to 181405)
TITLE      Waterston,R.
JOURNAL      Direct Submission
SUBMITTED      Submitted (09-MAY-2001) Department of Genetics, Washington
UNIVERSITY      University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
ON OCT 30, 2000 this sequence version replaced gi:7630982.
COMMENT
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
-----
Summary Statistics
Center project name: H_NH0417F21
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tatenio,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

#### VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-480A6; the clone sequenced to the right is RP11-536C12. Actual start of this clone is at base position 1 of RP11-417F21; actual end is at base position 181405 of RP11-417F21.

FEATURES	Location/Qualifiers
repeat_region	1. 181405
repeat_region	/organism="Homo sapiens"
repeat_region	/db_xref="taxon:9606"
repeat_region	/chromosome="2"
repeat_region	/map="2"
repeat_region	/clone="RP11-417F21"
repeat_region	/clone_lib="RPCI-11"
repeat_region	121..245
repeat_region	/rpt_family="Alu"
repeat_region	246..339
repeat_region	/rpt_family="MIR"
repeat_region	509..644
repeat_region	/rpt_family="Alu"
repeat_region	1040..1168
repeat_region	/rpt_family="L2"
repeat_region	1247..1431
repeat_region	/rpt_family="ERV1"
repeat_region	1747..1868
repeat_region	/rpt_family="ERV1"
repeat_region	1993..2071
repeat_region	/rpt_family="ERV1"
repeat_region	2081..2278
repeat_region	/rpt_family="MIR"
repeat_region	2587..2884
repeat_region	/rpt_family="Alu"
repeat_region	3159..3321
repeat_region	/rpt_family="Alu"
repeat_region	3381..3524
repeat_region	/rpt_family="L1"
repeat_region	3746..3793
repeat_region	/rpt_family="MIR"
repeat_region	5249..5525
repeat_region	/rpt_family="Alu"
repeat_region	5541..5717
repeat_region	/rpt_family="MaLR"
repeat_region	6039..6203
repeat_region	/rpt_family="MER53"

```

repeat_region 12576..12678
/rpt_family="MIR"
repeat_region 12772..12935
/rpt_family="MaLR"
repeat_region 12938..13372
/rpt_family="ERV1"
repeat_region 13396..13630
/rpt_family="MaLR"
misc_feature 13958..14095
/notes="similar to EST AA4311175 (NID:g2114883) zw71h09.rl"
misc_feature 14798..14937
/notes="similar to EST AA4311175 (NID:g2114883) zw71h09.rl"
misc_feature 14830..15130
/notes="similar to EST AI449682 (NID:g4291711) mr63b08.x1"
misc_feature 14881..15157
/notes="similar to EST AA432195 (NID:g2114583) zw71h09.sl"
misc_feature 14975..15179
/notes="similar to EST AA435548 (NID:g2140462) zt85h03.sl"
repeat_region 15509..15722
/rpt_family="CR1"
repeat_region 15729..15847
/rpt_family="MIR"
repeat_region 15864..16076
/rpt_family="MIR"
repeat_region 16086..16330
/rpt_family="CR1"
repeat_region 18176..18482
/rpt_family="Alu"
repeat_region 18522..18594
/rpt_family="MIR"
misc_feature 18618..18818
/notes="similar to EST AA435548 (NID:g2140462) zt85h03.sl"
misc_feature 18635..18818
/notes="similar to EST AI449682 (NID:g4291711) mr63b08.x1"
misc_feature 18637..18818
/notes="similar to EST AA432195 (NID:g2114583) zw71h09.sl"
repeat_region 19256..19566
/rpt_family="MaLR"
repeat_region 21258..21352
/rpt_family="MIR"
repeat_region 21921..22276
/rpt_family="MaLR"
repeat_region 23191..23287
/rpt_family="MER1_type"
repeat_region 23300..23447
/rpt_family="MER1_type"
repeat_region 23502..23782
/rpt_family="Alu"
repeat_region 24714..24847
/rpt_family="Achobo"
repeat_region 25831..26124
/rpt_family="MaLR"
repeat_region 26125..26250
/rpt_family="L1"
repeat_region 26251..26344
/rpt_family="MaLR"
repeat_region 26346..26422
/rpt_family="MIR"
repeat_region 26439..26504
/rpt_family="L2"
repeat_region 26669..26779
/rpt_family="Alu"
repeat_region 27377..27681
/rpt_family="Alu"
repeat_region 27699..27989
/rpt_family="MER1_type"
repeat_region 29138..29640
/rpt_family="MaLR"
repeat_region 30036..30460
/rpt_family="L1"
repeat_region 30581..30817
/rpt_family="Achobo"
repeat_region 30825..31103

```

```

repeat_region /rpt_family="Alu"
31257..31493
/rpt_family="Achobo"
31951..32103
/rpt_family="MER1_type"
32099..32198
/rpt_family="MER1_type"
32588..32692
/rpt_family="ERV1"

Query Match 2.5%; Score 21; DB 9; Length 181405;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 768 TCAAAACCACATCTGCTTCAGT 788
|||||
Db 23398 TCAAAACCACATCTGCTTCAGT 23378

RESULT 25
AP000871 199116 bp DNA linear PRI 11-JUL-2002
LOCUS Homo sapiens genomic DNA, chromosome 11 clone:RP11-709J7, complete
DEFINITION sequence.
ACCESSION AP000871
VERSION AP000871.5 GI:21728134
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:RP11-709J7.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens genomic DNA
JOURNAL Published Only in Database (1999)
REFERENCE 2 (bases 1 to 199116)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (13-DEC-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170
COMMENT On Jul 10, 2002 this sequence version replaced gi:21327919.
FEATURES
source
Location/Qualifiers
1..199116
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-709J7"

BASE COUNT 61419 a 40973 c 40167 g 56557 t
ORIGIN

Query Match 2.5%; Score 21; DB 9; Length 199116;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 767 CTCAAACCACATCTGCTTCAG 787
|||||
Db 152171 CTCAAACCACATCTGCTTCAG 152191

RESULT 26
AC102104 201306 bp DNA linear HTG 21-AUG-2002
LOCUS Mus musculus clone RP23-85P16, WORKING DRAFT SEQUENCE, 34 unordered
DEFINITION pieces.
ACCESSION AC102104
VERSION AC102104.2 GI:22381251
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

```

SOURCE  
ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS

1 (bases 1 to 201306)  
Birren, B., Nusbaum, C. and Lander, E.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Mus musculus, clone RP23-85P16  
Unpublished  
2 (bases 1 to 201306)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
Choepel, J., Colangelo, M., Collins, S., Collamore, A., Cook, A.,  
Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,  
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,  
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrim, J.,  
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D.,  
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Direct Submission  
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 201306)  
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,  
Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,  
Camarata, J., Chang, J., Chazaro, B., Choepel, J., Collamore, A.,  
Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,  
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,  
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,  
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,  
Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,  
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,  
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,  
O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,  
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P.,  
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,  
Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,  
Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 21, 2002 this sequence version replaced gi:17061190.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L18043  
Center clone name: 85\_P\_16  
----- Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 190693 bases at least Q40

Consensus quality: 195822 bases at least Q30  
Consensus quality: 197346 bases at least Q20  
Insert size: 178000; agarose-fp  
Insert size: 198006; sum-of-contigs  
Quality coverage: 10.2 in Q20 bases; agarose-fp  
Quality coverage: 9.1 in Q20 bases; sum-of-contigs  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 34 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 23629: contig of 23629 bp in length  
\* 23630 23729: gap of 100 bp  
\* 23730 24339: contig of 610 bp in length  
\* 24340 24439: gap of 100 bp  
\* 24440 25114: contig of 675 bp in length  
\* 25115 25214: gap of 100 bp  
\* 25215 26320: contig of 1106 bp in length  
\* 26321 26420: gap of 100 bp  
\* 26421 27032: contig of 612 bp in length  
\* 27033 27132: gap of 100 bp  
\* 27133 27644: contig of 517 bp in length  
\* 27650 27749: gap of 100 bp  
\* 27750 28451: contig of 702 bp in length  
\* 28452 28551: gap of 100 bp  
\* 28552 29365: contig of 814 bp in length  
\* 29366 29465: gap of 100 bp  
\* 29466 30503: contig of 1038 bp in length  
\* 30504 30603: gap of 100 bp  
\* 30604 31599: contig of 996 bp in length  
\* 31600 31699: gap of 100 bp  
\* 31700 32642: contig of 943 bp in length  
\* 32643 32742: gap of 100 bp  
\* 32743 33361: contig of 619 bp in length  
\* 33362 33461: gap of 100 bp  
\* 33462 34303: contig of 842 bp in length  
\* 34304 34403: gap of 100 bp  
\* 34404 35068: contig of 665 bp in length  
\* 35069 35168: gap of 100 bp  
\* 35169 36066: contig of 898 bp in length  
\* 36067 36166: gap of 100 bp  
\* 36167 37511: contig of 1345 bp in length  
\* 37512 37611: gap of 100 bp  
\* 37612 39137: contig of 1526 bp in length  
\* 39138 39237: gap of 100 bp  
\* 39238 40339: contig of 1102 bp in length  
\* 40340 40439: gap of 100 bp  
\* 40440 41651: contig of 1212 bp in length  
\* 41652 41751: gap of 100 bp  
\* 41752 42756: contig of 1005 bp in length  
\* 42757 42856: gap of 100 bp  
\* 42857 44691: contig of 1835 bp in length  
\* 44692 44791: gap of 100 bp  
\* 44792 46287: contig of 1496 bp in length  
\* 46288 46387: gap of 100 bp  
\* 46388 47554: contig of 1167 bp in length  
\* 47555 47654: gap of 100 bp  
\* 47655 49056: contig of 1402 bp in length  
\* 49057 49156: gap of 100 bp  
\* 49157 50534: contig of 1378 bp in length  
\* 50535 50634: gap of 100 bp  
\* 50635 52420: contig of 1786 bp in length  
\* 52421 52520: gap of 100 bp  
\* 52521 55044: contig of 2524 bp in length  
\* 55045 55144: gap of 100 bp  
\* 55145 56737: contig of 1593 bp in length  
\* 56738 56837: gap of 100 bp  
\* 56838 60234: contig of 3397 bp in length  
\* 60235 60334: gap of 100 bp

\* 60335 117900: contig of 57566 bp in length  
 \* 117901 118000: gap of 100 bp  
 \* 118001 129003: contig of 12903 bp in length  
 \* 130904 131003: gap of 100 bp  
 \* 131004 159946: contig of 28943 bp in length  
 \* 159947 160046: gap of 100 bp  
 \* 160047 197737: contig of 37691 bp in length  
 \* 197738 197837: gap of 100 bp  
 \* 197838 201306: contig of 3469 bp in length.

## FEATURES

source  
 1. .201306  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone\_lib="RP23-85P16"  
 /clone\_lib="RPCI-23 Female Mouse BAC"

## misc\_feature

1. .23629  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:left

## misc\_feature

23730. .24339  
 /note="assembly\_fragment"

## misc\_feature

24440. .25114

## misc\_feature

25215. .26320  
 /note="assembly\_fragment"

## misc\_feature

26421. .27032  
 /note="assembly\_fragment"

## misc\_feature

27133. .27649  
 /note="assembly\_fragment"

## misc\_feature

27750. .28451  
 /note="assembly\_fragment"

## misc\_feature

28552. .29365  
 /note="assembly\_fragment"

## misc\_feature

29466. .30503  
 /note="assembly\_fragment"

## misc\_feature

30604. .31599  
 /note="assembly\_fragment"

## misc\_feature

31700. .32642  
 /note="assembly\_fragment"

## misc\_feature

32743. .33361  
 /note="assembly\_fragment"

## misc\_feature

33462. .34303  
 /note="assembly\_fragment"

## misc\_feature

34404. .35068  
 /note="assembly\_fragment"

## misc\_feature

35169. .36066  
 /note="assembly\_fragment"

## misc\_feature

36167. .37511

Query Match 2.5%; Score 21; DB 2; Length 201306;

Best Local Similarity 100.0%; Pred. No. 5.4;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 710 TTGCTAATAAATTCAGAGA 730

Db 145913 TTGCTAATAAATTCAGAGA 145933

## RESULT 27

## AP002453

LOCUS AP002453 211344 bp DNA linear PRI 06-MAR-2001  
 DEFINITION Homo sapiens genomic DNA, chromosome 11q, clone:RP11-801G16,  
 complete sequence.

## ACCESSION

AP002453

## VERSION

AP002453.3 GI:13359385

## KEYWORDS

## HTG.

## SOURCE

## ORGANISM

Homo sapiens DNA, clone:RP11-801G16.  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

## 1 (sites)

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

## TITLE

## JOURNAL

Published Only in Database (2000) In press

## REFERENCE

2 (bases 1 to 211344)

## AUTHORS

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.  
 Direct Submission  
 Submitted (02-JUN-2000) Masahira Hattori, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 1-7-22 Suehiro-chou, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,  
 Tel:81-45-503-9111, Fax:81-45-503-9170)

## COMMENT

On Mar 16, 2001 this sequence version replaced gi:11559295.

## FEATURES

## source

Location/Qualifiers  
 1. .211344  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="11"  
 /map="11q"  
 /clone="RP11-801G16"

BASE COUNT 61435 a 41254 c 42802 g 65853 t

## ORIGIN

Query Match 2.5%; Score 21; DB 9; Length 211344;

Best Local Similarity 100.0%; Pred. No. 5.3;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 767 CTCAAACCATCTGCTTCAG 787

|||||

Db 32270 CTCAAACCATCTGCTTCAG 32290

## RESULT 28

## AC125236

## LOCUS

AC125236 14679 bp DNA linear PRI 01-AUG-2002

## DEFINITION

Homo sapiens BAC clone RP11-590J1 from 4, complete sequence.

## ACCESSION

AC125236

## VERSION

AC125236.4 GI:21954056

## KEYWORDS

## HTG.

## SOURCE

## human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 14679)  
 Sulston,J.E. and Waterston,R.  
 Toward a complete human genome sequence  
 Genome Res. 8 (11), 1097-1108 (1998)  
 95063792  
 2 (bases 1 to 14679)  
 Abbott,S. and Bielicki,L.  
 The sequence of Homo sapiens BAC clone RP11-590J1

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Jul 24, 2002 this sequence version replaced gi:21903696.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Web site: http://genome.wustl.edu/gsc  
 Contact: sapiens@watson.wustl.edu  
 ----- Summary Statistics  
 -----  
 Center project name: H\_NH0590J01

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenio, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
 VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-381K8, 2000 bp overlap; the clone sequenced to the right is RP11-427M12, 2000 bp overlap. Actual start of this clone is at base position 2291 of RP11-381K8; actual end is at base position 98258 of RP11-427M12.

Single plasmid region exists between 7383 and 7384.

#### FEATURES

##### Source

1. 14679  
 /location="Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosomes="4"  
 /map="4"  
 /clone="RP11-590J1"  
 /clone\_lib="RPCI-11"

repeat\_region  
 repeat\_region  
 repeat\_region  
 repeat\_region  
 repeat\_region  
 repeat\_region  
 repeat\_region  
 repeat\_region  
 repeat\_region  
 repeat\_region

repeat\_region 11008..11134  
 /rpt\_family="MaLR"  
 repeat\_region 11921..12060  
 /rpt\_family="MIR"  
 repeat\_region 13460..13584  
 /rpt\_family="AcHobo"  
 repeat\_region 13712..13808  
 /rpt\_family="L2"  
 repeat\_region 13848..14265  
 /rpt\_family="ERVL"  
 BASE COUNT 5215 a 2872 c 2696 g 3896 t  
 ORIGIN

Query Match 2.4%; Score 20; DB 9; Length 14679;  
 Best Local Similarity 100.0%; Pred. NO. 24;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 295 AACCTAATATCAGGATTTTC 314  
 |||||  
 Db 12198 AACCTAATATCAGGATTTTC 12217  
 |||||

#### RESULT 29

AC129525 40629 bp DNA linear HTG 30-JUL-2002  
 LOCUS Mus musculus clone RP23-72M4, LOW-PASS SEQUENCE SAMPLING.  
 DEFINITION AC129525  
 AC129525  
 VERSION AC129525.1 GI:22004339  
 HTG; HTGS PHASE0.  
 KEYWORDS house mouse.  
 SOURCE

#### ORGANISM

Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 40629)

#### REFERENCE

Birren, B., Nusbaum, C. and Lander, E.  
 Mus musculus chromosome, clone RP23-72M4  
 Unpublished

#### JOURNAL

2 (bases 1 to 40629)

#### REFERENCE

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Canarata, J., Chang, J., Chazaro, B., Choepel, Y., Collamore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliiev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kellis, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamae, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (30-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

#### TITLE

#### JOURNAL

#### COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: L14606  
 Center clone name: 72\_M\_4  
 -----

\* NOTE: This record contains 51 individual  
 \* sequencing reads that have not been assembled into

\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

\* 1 688: contig of 688 bp in length  
\* 789 788: gap of 100 bp  
\* 789 1499: contig of 711 bp in length  
\* 1500 1599: gap of 100 bp  
\* 1600 2287: contig of 688 bp in length  
\* 2288 2387: gap of 100 bp  
\* 2388 3075: contig of 688 bp in length  
\* 3076 3175: gap of 100 bp  
\* 3176 3861: contig of 686 bp in length  
\* 3862 3961: gap of 100 bp  
\* 3962 4665: contig of 704 bp in length  
\* 4666 4765: gap of 100 bp  
\* 4766 5479: contig of 714 bp in length  
\* 5480 5579: gap of 100 bp  
\* 5580 6297: contig of 718 bp in length  
\* 6298 6397: gap of 100 bp  
\* 6398 7096: contig of 699 bp in length  
\* 7097 7196: gap of 100 bp  
\* 7197 7858: contig of 662 bp in length  
\* 7859 7958: gap of 100 bp  
\* 7959 8658: contig of 700 bp in length  
\* 8659 8758: gap of 100 bp  
\* 8759 9462: contig of 704 bp in length  
\* 9463 9562: gap of 100 bp  
\* 9563 10283: contig of 721 bp in length  
\* 10284 10383: gap of 100 bp  
\* 10384 11094: contig of 711 bp in length  
\* 11095 11194: gap of 100 bp  
\* 11195 11894: contig of 700 bp in length  
\* 11895 11994: gap of 100 bp  
\* 11995 12697: contig of 703 bp in length  
\* 12698 12797: gap of 100 bp  
\* 12798 13502: contig of 705 bp in length  
\* 13503 13602: gap of 100 bp  
\* 13603 14322: contig of 720 bp in length  
\* 14323 14422: gap of 100 bp  
\* 14423 15112: contig of 690 bp in length  
\* 15113 15212: gap of 100 bp  
\* 15213 15904: contig of 692 bp in length  
\* 15905 16004: gap of 100 bp  
\* 16005 16716: contig of 712 bp in length  
\* 16717 16816: gap of 100 bp  
\* 16817 17416: contig of 600 bp in length  
\* 17417 17516: gap of 100 bp  
\* 17517 18238: contig of 722 bp in length  
\* 18239 18338: gap of 100 bp  
\* 18339 18942: contig of 604 bp in length  
\* 18943 19042: gap of 100 bp  
\* 19043 19727: contig of 685 bp in length  
\* 19728 19827: gap of 100 bp  
\* 19828 20520: contig of 693 bp in length  
\* 20521 20620: gap of 100 bp  
\* 20621 21331: contig of 711 bp in length  
\* 21332 21431: gap of 100 bp  
\* 21432 22158: contig of 727 bp in length  
\* 22159 22258: gap of 100 bp  
\* 22259 22934: contig of 676 bp in length  
\* 22935 23034: gap of 100 bp  
\* 23035 23732: contig of 698 bp in length  
\* 23733 23832: gap of 100 bp  
\* 23833 24527: contig of 695 bp in length  
\* 24528 24627: gap of 100 bp  
\* 24628 25322: contig of 695 bp in length  
\* 25323 25422: gap of 100 bp

\* 25423 26135: contig of 713 bp in length  
\* 26136 26235: gap of 100 bp  
\* 26236 26913: contig of 678 bp in length  
\* 26914 27013: gap of 100 bp  
\* 27014 27729: contig of 716 bp in length  
\* 27730 27829: gap of 100 bp  
\* 27830 28554: contig of 725 bp in length  
\* 28555 28654: gap of 100 bp  
\* 28655 29338: contig of 684 bp in length  
\* 29339 29438: gap of 100 bp  
\* 29439 30141: contig of 703 bp in length  
\* 30142 30241: gap of 100 bp  
\* 30242 30942: contig of 701 bp in length  
\* 30943 31042: gap of 100 bp  
\* 31043 31756: contig of 714 bp in length  
\* 31757 31856: gap of 100 bp  
\* 31857 32566: contig of 710 bp in length  
\* 32567 32666: gap of 100 bp  
\* 32667 33354: contig of 688 bp in length  
\* 33355 33454: gap of 100 bp  
\* 33455 34179: contig of 725 bp in length  
\* 34180 34279: gap of 100 bp  
\* 34280 34996: contig of 717 bp in length  
\* 34997 35086: gap of 100 bp  
\* 35097 35802: contig of 706 bp in length  
\* 35803 35902: gap of 100 bp  
\* 35903 36589: contig of 687 bp in length  
\* 36590 36689: gap of 100 bp  
\* 36690 37387: contig of 698 bp in length  
\* 37388 37487: gap of 100 bp  
\* 37488 38186: contig of 699 bp in length  
\* 38187 38286: gap of 100 bp  
\* 38287 39000: contig of 714 bp in length  
\* 39001 39100: gap of 100 bp  
\* 39101 39823: contig of 723 bp in length  
\* 39824 39923: gap of 100 bp  
\* 39924 40629: contig of 706 bp in length.

FEATURES  
source

1. 40629  
/organism="Mus musculus"  
/db xref="taxon:10090"  
/clone="RP23-72M4"  
/clone\_lib="RPCI-23 Female Mouse BAC"  
BASE COUNT 10467 a 7141 c 7211 g 10750 t 5060 others  
ORIGIN

Query Match 2.48: Score 20; DB 2; Length 40629;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 268 GTAGCTCCAGGCATTGATT 287  
|||||  
Db 29858 GTAGCTCCAGGCATTGATT 29877

RESULT 30

AC115981  
LOCUS AC115981 60058 bp DNA linear HTG 22-MAR-2002  
DEFINITION Mus musculus clone RP24-140N9, LOW-PASS SEQUENCE SAMPLING.  
ACCESSION AC115981  
VERSION AC115981.1 GI:19683660  
KEYWORDS HTG; HTGS PHASE0.  
SOURCE Mus musculus.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 60058)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Mus musculus, clone RP24-140N9  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 60058)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,

Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,  
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,  
 Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
 Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, J.,  
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kanat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,  
 Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,  
 MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
 Michovita, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L.,  
 Mihaeva, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,  
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D.,  
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
 Raymond, C., Retta, R., Ribbeck, M., Riley, R., Rise, C., Rogov, P.,  
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

# TITLE

## JOURNAL

Direct Submission

Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L25087

Center clone name: 140\_N\_9

-----

\* NOTE: This record contains 77 individual

\* sequencing reads that have not been assembled into

\* contigs. Runs of N are used to separate the reads

\* and the order in which they appear is completely

\* arbitrary. Low-pass sequence sampling is useful for

\* identifying clones that may be gene-rich and allows

\* overlap relationships among clones to be deduced.

\* However, it should not be assumed that this clone

\* will be sequenced to completion. In the event that

\* the record is updated, the accession number will

\* be preserved.

\* 1 686: contig of 686 bp in length

\* 687 786: gap of 100 bp

\* 787 1471: contig of 685 bp in length

\* 1472 1571: gap of 100 bp

\* 1572 2190: contig of 619 bp in length

\* 2191 2290: gap of 100 bp

\* 2291 2966: contig of 676 bp in length

\* 2967 3066: gap of 100 bp

\* 3067 3752: contig of 685 bp in length

\* 3753 3852: gap of 100 bp

\* 3853 4540: contig of 688 bp in length

\* 4541 4640: gap of 100 bp

\* 4641 5331: contig of 691 bp in length

\* 5332 5431: gap of 100 bp

\* 5432 6124: contig of 693 bp in length

\* 6125 6224: gap of 100 bp

\* 6225 6901: contig of 677 bp in length

\* 6902 7001: gap of 100 bp

\* 7002 7688: contig of 687 bp in length

\* 7689 7788: gap of 100 bp

\* 7789 8361: contig of 573 bp in length

\* 8362 8461: gap of 100 bp

\* 8462 9119: contig of 658 bp in length

\* 9120 9219: gap of 100 bp

\* 9220 9902: contig of 683 bp in length

\* 9903 10002: gap of 100 bp

\* 10003 10694: contig of 692 bp in length

\* 10695 10794: gap of 100 bp

\* 10795 11490: contig of 696 bp in length

\* 11491 11590: gap of 100 bp

\* 11591 12288: contig of 692 bp in length

\* 12283 12382: gap of 100 bp

\* 12383 13058: contig of 676 bp in length

\* 13059 13158: gap of 100 bp

\* 13159 13817: contig of 659 bp in length

\* 13818 13917: gap of 100 bp

\* 13918 14606: contig of 689 bp in length

\* 14607 14706: gap of 100 bp

\* 14707 15368: contig of 662 bp in length

\* 15369 15468: gap of 100 bp

\* 15469 16155: contig of 687 bp in length

\* 16156 16255: gap of 100 bp

\* 16256 16958: contig of 703 bp in length

\* 16959 17058: gap of 100 bp

\* 17059 17742: contig of 684 bp in length

\* 17743 17842: gap of 100 bp

\* 17843 18470: contig of 628 bp in length

\* 18471 18570: gap of 100 bp

\* 18571 19259: contig of 689 bp in length

\* 19260 19359: gap of 100 bp

\* 19360 20042: contig of 683 bp in length

\* 20043 20142: gap of 100 bp

\* 20143 20833: contig of 691 bp in length

\* 20834 20933: gap of 100 bp

\* 20934 21625: contig of 692 bp in length

\* 21626 21725: gap of 100 bp

\* 21726 22410: contig of 685 bp in length

\* 22411 22510: gap of 100 bp

\* 22511 23198: contig of 688 bp in length

\* 23199 23298: gap of 100 bp

\* 23299 23982: contig of 684 bp in length

\* 23983 24082: gap of 100 bp

\* 24083 24772: contig of 690 bp in length

\* 24773 24872: gap of 100 bp

\* 24873 25558: contig of 686 bp in length

\* 25559 25658: gap of 100 bp

\* 25659 26363: contig of 705 bp in length

\* 26364 26463: gap of 100 bp

\* 26464 27156: contig of 693 bp in length

\* 27157 27256: gap of 100 bp

\* 27257 27953: contig of 697 bp in length

\* 27954 28053: gap of 100 bp

\* 28054 28738: contig of 685 bp in length

\* 28739 28838: gap of 100 bp

\* 28839 29526: contig of 688 bp in length

\* 29527 29626: gap of 100 bp

\* 29627 30316: contig of 630 bp in length

\* 30317 30416: gap of 100 bp

\* 30417 31103: contig of 687 bp in length

\* 31104 31203: gap of 100 bp

\* 31204 31892: contig of 689 bp in length

\* 31893 31992: gap of 100 bp

\* 31993 32691: contig of 699 bp in length

\* 32692 32791: gap of 100 bp

\* 32792 33477: contig of 686 bp in length

\* 33478 33577: gap of 100 bp

\* 33578 34263: contig of 686 bp in length

\* 34264 34363: gap of 100 bp

\* 34364 35045: contig of 682 bp in length

\* 35046 35145: gap of 100 bp

\* 35146 35754: contig of 609 bp in length

\* 35755 35854: gap of 100 bp

\* 35855 36509: contig of 655 bp in length

\* 36510 36609: gap of 100 bp

\* 36610 37293: contig of 684 bp in length

\* 37294 37393: gap of 100 bp

\* 37394 38081: contig of 688 bp in length

\* 38082 38181: gap of 100 bp

\* 38182 38832: contig of 651 bp in length

\* 38833 38932: gap of 100 bp



```

* 38933 33602: contig of 670 bp in length
* 39603 39702: gap of 100 bp
* 39703 40398: contig of 696 bp in length
* 40399 40498: gap of 100 bp
* 40499 41184: contig of 686 bp in length
* 41185 41284: gap of 100 bp
* 41285 41968: contig of 684 bp in length
* 41969 42068: gap of 100 bp
* 42069 42731: contig of 663 bp in length
* 42732 42831: gap of 100 bp
* 42832 43537: contig of 706 bp in length
* 43538 43637: gap of 100 bp
* 43638 44332: contig of 695 bp in length
* 44333 44432: gap of 100 bp
* 44433 45112: contig of 680 bp in length
* 45113 45212: gap of 100 bp
* 45213 45886: contig of 674 bp in length
* 45887 45986: gap of 100 bp
* 45987 46675: contig of 689 bp in length
* 46676 46775: gap of 100 bp
* 46776 47439: contig of 664 bp in length
* 47440 47539: gap of 100 bp
* 47540 48231: contig of 692 bp in length
* 48232 48331: gap of 100 bp
* 48332 49013: contig of 682 bp in length
* 49014 49113: gap of 100 bp
* 49114 49803: contig of 690 bp in length
* 49804 49903: gap of 100 bp
* 49904 50579: contig of 676 bp in length
* 50580 50678: gap of 100 bp
* 50680 51372: contig of 693 bp in length
* 51373 51472: gap of 100 bp
* 51473 52163: contig of 691 bp in length
* 52164 52263: gap of 100 bp
* 52264 52956: contig of 693 bp in length
* 52957 53056: gap of 100 bp
* 53057 53574: contig of 686 bp in length
* 53743 53842: gap of 100 bp

```

Query Match 2.4%; Score 20; DB 2; Length 60058;

Best Local Similarity 100.0%; Pred. No. 21;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 268 GTAGCTCAGGCATGATT 287

Db 16130 GTAGCTCAGGCATGATT 16149

RESULT 31  
AC004996  
LOCUS Homo sapiens PAC clone RP5-1200123 from 7p15, complete sequence.  
DEFINITION  
AC004996  
VERSION  
AC004996.1 GI:3482962  
KEYWORDS  
HTG.

SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 100921)  
Latreille,P., Elliott,G. and Biewald,T.  
The sequence of Homo sapiens PAC clone RP5-1200123  
Unpublished  
2 (bases 1 to 100921)  
Waterston,R.H.  
Direct Submission  
Submitted (12-JUN-1998) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
3 (bases 1 to 100921)  
Waterston,R.  
Direct Submission  
Submitted (28-AUG-1998) Department of Genetics, Washington

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
4 (bases 1 to 100921)  
Waterston,R.  
Direct Submission  
Submitted (21-DEC-1999) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Aug 28, 1998 this sequence version replaced gi:3213001.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc  
Contact: sapiens@watson.wustl.edu  
----- Summary Statistics  
-----  
Center project name: H\_DJ1200123

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/OTB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at Roswell Park Cancer Institute, using the method described by Ioannou et al., Nature Genetics, 6:84-9 (1994). The library is from one male donor. For further details, see <http://bacpac.med.buffalo.edu/>. The clone is available from Genome Systems, Inc. (<http://www.genomesystems.com>).

#### VECTOR: PCYPAC2

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP1-170019, 200 bp overlap; the clone sequenced to the right is RP5-110315, 200 bp overlap. Actual start of this clone is at base position 80816 of RP1-170019; actual end is at 100725 of RP5-1200123.

Bases 25817 to 28047 are the result of sequence obtained by PCR of genomic DNA from PAC RP5-1200123.

FEATURES  
source  
Location/Qualifiers  
1. 100921  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7p15"  
/clone="RP5-1200123"  
/clone\_lib="RPCI-5"  
30. 162  
/rpt\_family="Alu"  
repeat\_region  
266. 566  
/rpt\_family="Alu"  
repeat\_region  
587. 638  
/rpt\_family="MIR"  
repeat\_region  
857. 1140  
/rpt\_family="Alu"  
repeat\_region  
1681. 1974  
/rpt\_family="Alu"

```

repeat_region 2750..2906
/rpt_family="MER1_type"
repeat_region 3149..3453
/rpt_family="Alu"
repeat_region 3686..3875
/rpt_family="Alu"
repeat_region 3918..4022
/rpt_family="Alu"
repeat_region 4145..4445
/rpt_family="Alu"
repeat_region 5774..5932
/rpt_family="Alu"
repeat_region 6828..7128
/rpt_family="Alu"
repeat_region 7281..7578
/rpt_family="Alu"
repeat_region 8430..8732
/rpt_family="Alu"
repeat_region 9536..9837
/rpt_family="Alu"
repeat_region 9893..10429
/rpt_family="L1"
repeat_region 10446..10753
/rpt_family="Alu"
repeat_region 10760..10829
/rpt_family="L1"
repeat_region 10794..10839
/rpt_family="L1"
repeat_region 10885..11074
/rpt_family="L1"
repeat_region 11075..11368
/rpt_family="Alu"
repeat_region 11369..11648
/rpt_family="L1"
repeat_region 11672..11970
/rpt_family="Alu"
repeat_region 11971..12105
/rpt_family="Alu"
repeat_region 12163..12464
/rpt_family="Alu"
repeat_region 12495..12791
/rpt_family="Alu"
repeat_region 12792..13518
/rpt_family="L1"
misc_feature 13532..13933
/note="similar to EST AA489323 (NID:G2218925) ab36g09.s1"
misc_feature complement(13564..13927)
/note="similar to Mus musculus EST AAG70449 (NID:G2965894)
vq21g07.r1"
misc_feature complement(13568..13926)
/note="60S ribosomal protein pseudogene"
repeat_region 13974..14267
/rpt_family="L1"
repeat_region 14278..14576
/rpt_family="Alu"
repeat_region 14587..14874
/rpt_family="Alu"
repeat_region 14880..15374
/rpt_family="L1"
repeat_region 15386..15655
/rpt_family="Alu"
repeat_region 15661..16369
/rpt_family="L1"
repeat_region 16585..16657
/rpt_family="L2"
repeat_region 16796..17041
/rpt_family="L1"
repeat_region 17079..17376
/rpt_family="Alu"
repeat_region 17397..17441
/rpt_family="Alu"
repeat_region 17442..17730
/rpt_family="Alu"

```

```

repeat_region 17879..17940
/rpt_family="L2"
repeat_region 17933..18372
/rpt_family="L2"
repeat_region 18426..18592
/rpt_family="Alu"
repeat_region 18830..19110
/rpt_family="L2"
repeat_region 19285..19424
/rpt_family="MIR"
repeat_region 19632..19931
/rpt_family="Alu"
repeat_region 19991..20030
/rpt_family="MIR"
repeat_region 20162..20305
/rpt_family="Alu"
repeat_region 20306..21086
/rpt_family="Retroviral"
repeat_region 21087..21232
/rpt_family="Alu"
repeat_region 21480..21781
/rpt_family="Alu"
misc_feature 21761..22150
/note="similar to EST AA420757 (NID:G2094636) nc63e05.s1"
misc_feature complement(21858..22352)
/note="match to EST AA420798 (NID:G2094704) nc63e05.r1"
repeat_region 22441..22743
/rpt_family="Alu"
repeat_region 22844..22979
/rpt_family="Alu"
repeat_region 22980..23276

Query Match 2.4%; Score 20; DB 9; Length 100921;
Best Local Similarity 100.0%; Pred.No.20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 616 AATCTAAATTTGCTTACCA 635
|||||
Db 64898 AATCTAAATTTGCTTACCA 64917

RESULT 32
AP003815 106253 bp DNA linear PLN 03-AUG-2002
LOCUS Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 7,
DEFINITION BAC clone:OJ1163_G04.
ACCESSION AP003815
VERSION AP003815.2 GI:20269312
KEYWORDS Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
SOURCE clone:OJ1163_G04.
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
clone:OJ1163_G04
JOURNAL Published Only in Database (2001)
REFERENCE 2
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT On Apr 22, 2002 this sequence version replaced gi:14595155.
Genes were predicted from the integrated results of the following:
GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
(http://www.softberry.com/), GeneMark.hmm
(http://opal.biology.gatech.edu/GeneMark/), GlimmerM

```

([http://www.tigr.org/tdb/glimmer/glmr\\_form.html](http://www.tigr.org/tdb/glimmer/glmr_form.html)), RiceHMM (<http://rgp.dna.affrc.go.jp/RiceHMM/>), SplicePredictor (<http://bioinformatics.iastate.edu/cgi-bin/sp.cgi>), BLASTN and BLASTX. The genomic sequence was searched against NCBI NonRedundant Protein database, nr (<ftp://ncbi.nlm.nih.gov/blast/db/>) and the cDNA sequence database at RGP. Protein homologies of the coding regions were searched against NCBI NonRedundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DBJ accession no. and RGP clone ID.

A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence.

The orientation of the sequence is from M13rev to -21M13 of the BAC clone. This sequence of OJ1163 G04 clone has an overlap with OJ1282 E02 clone (DBJ: AP003950) at 5' end and an overlap with P0409B11 clone (DBJ: AP005185) at 3' end. The nucleotide sequence of this BAC clone was generated by combining Monsanto and RGP-Japan sequencing data. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rgp.dna.affrc.go.jp/GenomeSeq.html>.

Location/Qualifiers

1. 106253

/organism="Oryza sativa (japonica cultivar-group)"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/chromosome="7"

/clone="OJ1163\_G04"

3860. 4528

/gene="OJ1163\_G04.1"

3860. 4528

/gene="OJ1163\_G04.1"

/note="This category is not included in IRGSP standard. hypothetical ORF

predicted by GlimmerW"

join(7544. 7608,7878. 8616,8694. 8742,8826. 8890, 9000. 9054,10333. 10451,10572. 10638,11542. 11663, 11747. 12016)

/gene="OJ1163\_G04.2"

join(7544. 7608,7878. 8616,8694. 8742,8826. 8890, 9000. 9054,10333. 10451,10572. 10638,11542. 11663, 11747. 12016)

/gene="OJ1163\_G04.2"

/note="kinase-like protein"

/codon\_start=1

/protein\_id="BAC06968.1"

/db\_xref="GI:22093674"

/translation="MALAQPTLTQRCGPARYGESLPSESSPEKKLKLFPVKTK KYLEPNTMTQACKRRAVYISSESDSGTDEVEGSKLSDGVSTVTCGQPTS KNVDPMNTSKQSGILKLMKDKSGWIFNTEVDVYVIGDIPDVFVIRPMDLGTW KRKLTQYSNPYFAADVLRTFSNAMPKPPNDVHGIDQLNKIPDSWKLLERGV KRNLVQEPSLKVLAQAVTPKPLPKGVTAQNSAVSKTLATALSSKKVKIFSVR GSELTSSKDTPLOAVGRDCTINGSLPCTKDNAKTPRQSEDSSESTEYLPPEPL SPFKALRAAMLKSPFAGTIVQAQKALDDGKKIDPAKLQERLERKQOEKARIE AQVKAABAAQKLDEEMRMKQERARALAHMKKTVIDINDSDFKOLENLSKKW ELNPPGKLIIVDFDGLPGLGSLERLGLFMKCOLKEEVEHEMEDSDSPSTEIDVE EGETISFCQ"

complement(join(13799. 15368,15484. 15636,15731. 15792, 16081. 16329,16679. 16858,16939. 17168,17784. 17961, 18553. 18658,19012. 19109,19509. 20039,20134. 20313, 20405. 20669,20783. 20853,20933. 21148,21373. 21493, 21588. 21593))

/gene="OJ1163\_G04.3"

complement(join(13799. 15368,15484. 15636,15731. 15792, 16081. 16329,16679. 16858,16939. 17168,17784. 17961, 18553. 18658,19012. 19109,19509. 20039,20134. 20313,

20405. 20669,20783. 20853,20933. 21148,21373. 21493, 21588. 21593))

/gene="OJ1163\_G04.3"

/codon\_start=1

/product="putative pre-mRNA splicing factor ATP-dependent RNA helicase"

/protein\_id="BAC06969.1"

/db\_xref="GI:22093675"

/translation="MEGDMATMTLTGPEDDTGAQGLILFSRDRVMYRPPFGKSALGL DLAAHRAEGGNFAKPPQKVVAAATSIDDEKPGPAESDEKSLSGHRGSVRRY RGANSRDTSEKPTITDEGRGSPSHRDSYRODTHKSRSSOGSHRSRTPRYDDY DLAKRREAGGNFAKPPQKVVAAATSIDDEKPGPAESDEKSLSGHRGSVRRY SSSSRTPAKSDMDSDGKWEEDTPRREIRDRSNHRHDERERSTSDVYNNKRSHH WLGENTPRYAASPDNDVNSPAPIRASGSGSGSYPSRSGSHOLTIFSSASANDREA DRSPDADNGYEISEEMQMENDYADWDCEHNTMFDGNSMYLEDSSYSKDRRA QLPKRLTRKDSGLMTLAQSKLSQADNAQMEDROLRLRGLVARGSVQTEFRDEDER KLVLLVHDTKPPPLDGRVFTKQAEPMPLKDPDTSMAIIVARKGSALVREIRKQSMN KSQRPFELAGSKLNLGVEKTAQVDATATVGDQGEIDPFKEAFESQHMVKYKAE VSPFAGKSUSQOQYLPIFTVRDDLQVRENVVVVGTGSGTKTQLTQYLHEDG YTTTIGVGTQPRRVAAMSVAKRSEMETELGHKGYAIRFEDMTSPNTIKYMTDG VLLRETLKADLDKRYIVMDEAHERSLNDVLFGILKVVARRDFKLIVTSATLNA DKFSKFGGTFPVNFSPKPCDYVEAAVKQAMTHITSGPGDILIFMTQGEIEAT CYALARELEOLISSSTKTPKLSLTIPIYSOLPADLOAKIFQKAEGRKTCIVATNIAE TSLTVDGIFVITDYGKMKYVPRMGMDALQVPSRAAADQQRAGRAGRTGPTCYR LFTSATQNEMLPNPVEIORTNLGNVLLSLKLVENLDQDFMDPPQENILNLSMT QLVGLGALNNVGALTIVGWKVEFDFDPTLAKMLLMEQLECLDEVLTIVISLSPVS FFRKORAEESDAAREKFFVPESDHLTLNVLQWKSNQYRGWDCNDHFLHVGILKRA REVSOLLDLIKTLKILPLTSCHEMDVVRKAI CSAYPHNAARLKGVEYVNCNRMGPC HLHPSALYGLGYTPDYYVYHELVTTLTKYMQCVTA VDPQWLAEGLGPMFVSGTMDTS LLHKKRQKEDKTAMEEEMKLRQEQAEARLEKEREKRAKQOQOQVSNPGLKGGST YLRPKRMGL"

join(23653. 23728,24290. 24723)

/gene="OJ1163\_G04.4"

join(23653. 23728,24290. 24723)

/gene="OJ1163\_G04.4"

/note="hypothetical protein predicted by GeneMark.hmm etc."

/codon\_start=1

/protein\_id="BAC06970.1"

/db\_xref="GI:22093676"

/translation="MGDNTNEKIMVQHNLTSRPSTQDPTTSSSSSGASWPELRR KWLPPDDKAREGAPNPPRAGRSLSRSELRLSLRVLHLVSPASNAHPRVASVLPS SLPRAFGAMAVATWWTTTAAVCNATIASLYLPVVCALATPTTLPLADAGERHHA KJHRQHISN"

join(27540. 28037,30442. 30465)

/gene="OJ1163\_G04.5"

join(27540. 28037,30442. 30465)

/gene="OJ1163\_G04.5"

/note="contains EST D23772 (R0119) unknown protein"

/codon\_start=1

/protein\_id="BAC06971.1"

/db\_xref="GI:22093677"

/translation="MAGRHRNPPPPPPRGGGVGRGHPPPPPSLHHLRLLPPPHLDD DFRPELLPPLHRLDDFRPHHLLPPPHHHHLLDFEPGPHHVFAGGGGAGPPAPH VAAALREERGAETIEARHLVQGRLAATHALVQEVSAVRHLELGTARGLAAQEEG ELRLREVQCSIE"

join(30675. 30892,31279. 31433,31646. 31722,31798. 31905)

/gene="OJ1163\_G04.6"

join(30675. 30892,31279. 31433,31646. 31722,31798. 31905)

/gene="OJ1163\_G04.6"

/note="similar to senescence-associated protein"

/codon\_start=1

/protein\_id="BAC06972.1"

/db\_xref="GI:22093678"

/translation="MKMEALRAVEENRAELAQVRLDIQKLAARQELMQGVQGFQD LARSAVLQQAALKAIEIQIRHETQHLRSGYAEVSGQGMQKGLISVAS EYFKLRAEAKGAAVSGGNQYVGGYGNPKAAYANPYNAGTINQHPQANTAS GSQFGPGSTHAPWGAYDMQATGRR"

complement(join(32914. 33030,33407. 33481,33563. 33652, 33732. 33794,33885. 34082,34199. 34300,35502. 35650, 36362. 36431,36686. 36808,37767. 37826))

/gene="OJ1163\_G04.7"



```

RESULT 35
AL450424
LOCUS
DEFINITION Homo sapiens chromosome 1 clone RP11-563I16, *** SEQUENCING IN
PROGRESS ***, 25 unordered pieces.
ACCESSION AL450424
VERSION AL450424.3 GI:12331147
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Mclay, K.
Direct Submission
Submitted (11-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jan 22, 2001 this sequence version replaced gi:11493361.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA563I16
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 13523 bases at least Q40
Consensus quality: 139636 bases at least Q30
Consensus quality: 141952 bases at least Q20
Insert size: 144405; sum-of-contigs
Insert size: 171032; 2.2% error; agarose-fp
Quality coverage: 3.60x in Q20 bases; sum-of-contigs Quality
coverage: 3.25x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 5105: contig of 5105 bp in length
* 5106 5205: gap of 100 bp
* 5206 10570: contig of 5365 bp in length
* 10571 10670: gap of 100 bp
* 10671 17626: contig of 6956 bp in length
* 17627 17726: gap of 100 bp
* 17727 20478: contig of 2752 bp in length
* 20479 20578: gap of 100 bp
* 20579 23035: contig of 2457 bp in length
* 23036 23135: gap of 100 bp
* 23136 25308: contig of 2173 bp in length
* 25309 25408: gap of 100 bp
* 25409 30469: contig of 5061 bp in length
* 30470 30569: gap of 100 bp
* 30570 32688: contig of 2119 bp in length
* 32689 32789: gap of 100 bp
* 32789 36187: contig of 3399 bp in length
* 36188 36287: gap of 100 bp
* 36288 48988: contig of 12701 bp in length
* 48989 49088: gap of 100 bp
* 49089 67333: contig of 18245 bp in length
* 67334 67433: gap of 100 bp
* 67434 82317: contig of 14884 bp in length
* 82318 82417: gap of 100 bp
* 82418 84614: contig of 2197 bp in length

```

```

* 84615 84714: gap of 100 bp
* 84715 90326: contig of 5612 bp in length
* 90327 90426: gap of 100 bp
* 90427 92647: contig of 2221 bp in length
* 92648 92747: gap of 100 bp
* 92748 97690: contig of 4943 bp in length
* 97691 97790: gap of 100 bp
* 97791 100799: contig of 3009 bp in length
* 100800 100899: gap of 100 bp
* 100900 104202: contig of 3303 bp in length
* 104203 104302: gap of 100 bp
* 104303 107157: contig of 2855 bp in length
* 107158 107257: gap of 100 bp
* 107258 110300: contig of 3043 bp in length
* 110301 110400: gap of 100 bp
* 110401 113955: contig of 3555 bp in length
* 113956 114055: gap of 100 bp
* 114056 118801: contig of 4746 bp in length
* 118802 118901: gap of 100 bp
* 118902 129831: contig of 10930 bp in length
* 129832 129931: gap of 100 bp
* 129932 137545: contig of 7614 bp in length
* 137546 137645: gap of 100 bp
* 137646 146805: contig of 9160 bp in length.

```

```

FEATURES
      source
        1..146805
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="1"
            /clone="RP11-563I16"
            /clone_lib="RPC1-11.2"
        1..5105
            /notes="assembly_fragment:00368"
            clone_end:T7
            vector_side:left
        5206..10570
            /notes="assembly_fragment:00594"
            fragment_chain:1
        10671..17626
            /notes="assembly_fragment:00786"
            fragment_chain:1
        17727..20478
            /notes="assembly_fragment:00844"
            fragment_chain:1
        20579..23035
            /notes="assembly_fragment:00011"
            fragment_chain:1
        23136..25308
            /notes="assembly_fragment:01029"
            fragment_chain:1
        25409..30469
            /notes="assembly_fragment:00148"
            fragment_chain:2
        30570..32688
            /notes="assembly_fragment:00961"
            fragment_chain:2
        32789..36187
            /notes="assembly_fragment:00310"
            fragment_chain:3
        36288..48988
            /notes="assembly_fragment:01150"
            fragment_chain:3
        49089..67333
            /notes="assembly_fragment:00444"
            fragment_chain:4
        67434..82317
            /notes="assembly_fragment:00140"
            fragment_chain:4
        82418..84614
            /notes="assembly_fragment:00657"
            fragment_chain:5
        84715..90326
            /note="assembly_fragment:00757"

```

```

misc_feature      fragment chain:5"
90427. .92847
/note="assembly fragment:00677
fragment chain:6"
92748. .97690
/note="assembly fragment:00248
fragment chain:6"
97791. .100799
/note="assembly fragment:00091"
100900. .104202
/note="assembly fragment:00538"
104303. .107157
/note="assembly fragment:00675"
107258. .110300
/note="assembly fragment:00695"
110401. .113955
/note="assembly fragment:01316"
114056. .118801
/note="assembly fragment:01292
fragment chain:7"
118902. .129831
/note="assembly fragment:00551
fragment chain:7"
129932. .137545
/note="assembly fragment:01410
fragment chain:7"
137646. .146805
/note="assembly fragment:00763
fragment chain:7
clone end:SP6
vector side:right"
BASE COUNT      37508 a 33240 c 34076 g 39557 t 2424 others
ORIGIN
Query Match      2.4%; Score 20; DB 2; Length 146805;
Best Local Similarity 100.0%; Pred. No.19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 614 TCAATCTAAATTTGCTTAC 633
Db 11445 TCAATCTAAATTTGCTTAC 11464

RESULT 36
AC021462/c
LOCUS
DEFINITION      Homo sapiens clone RP11-276C1, WORKING DRAFT SEQUENCE, 21 unordered
pieces.
ACCESSION      AC021462
VERSION      AC021462.3 GI:7387343
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens
ORGANISM      Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 153023)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-276C1
Unpublished
REFERENCE      2 (bases 1 to 153023)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karas,A., Klein,J.,
Lander,E., Lehotzky,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,

```

```

Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tefaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 3, 2000 this sequence version replaced gi:7230200.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5201
Center clone name: 276 C.1
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 141815 bases at least Q40
Consensus quality: 147368 bases at least Q30
Consensus quality: 149481 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 151023; sum-of-contigs
Quality coverage: 4.3 in Q20 bases; agarose-fp
Quality coverage: 4.5 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1015: contig of 1015 bp in length
* 1016 1115: gap of 100 bp
* 1116 2276: contig of 1161 bp in length
* 2277 2376: gap of 100 bp
* 2377 4009: contig of 1633 bp in length
* 4010 4109: gap of 100 bp
* 4110 5737: contig of 1628 bp in length
* 5738 5837: gap of 100 bp
* 5838 6939: contig of 1102 bp in length
* 6940 7039: gap of 100 bp
* 7040 9504: contig of 2465 bp in length
* 9505 9604: gap of 100 bp
* 9605 11734: contig of 2130 bp in length
* 11735 11834: gap of 100 bp
* 11835 14111: contig of 2277 bp in length
* 14112 14211: gap of 100 bp
* 14212 16973: contig of 2762 bp in length
* 16974 17073: gap of 100 bp
* 17074 19588: contig of 2515 bp in length
* 19589 19688: gap of 100 bp
* 19689 22275: contig of 2587 bp in length
* 22276 22375: gap of 100 bp
* 22376 25823: contig of 3448 bp in length
* 25824 25923: gap of 100 bp
* 25924 31307: contig of 5384 bp in length
* 31308 31407: gap of 100 bp
* 31408 34863: contig of 3456 bp in length
* 34864 34963: gap of 100 bp
* 34964 41382: contig of 6419 bp in length
* 41383 41482: gap of 100 bp
* 41483 49024: contig of 7542 bp in length
* 49025 49124: gap of 100 bp
* 49125 58672: contig of 9548 bp in length
* 58673 58772: gap of 100 bp

```

\* 58773 69622: contig of 10850 bp in length  
 \* 69623 69722: gap of 100 bp  
 \* 69723 88191: contig of 18469 bp in length  
 \* 88192 88291: gap of 100 bp  
 \* 88292 107084: contig of 18793 bp in length  
 \* 107085 107184: gap of 100 bp  
 \* 107185 153023: contig of 45839 bp in length.

## FEATURES

source  
 1. .153023  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="RP11-276C1"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 1. .1015  
 /note="assembly\_fragment"  
 1116. .2276  
 /note="assembly\_fragment"  
 2377. 4009  
 /note="assembly\_fragment"  
 4110. .5737  
 /note="assembly\_fragment"  
 5838. .6939  
 /note="assembly\_fragment"  
 7040. 9504  
 /note="assembly\_fragment"  
 9605. .11734  
 /note="assembly\_fragment"  
 11835. .14111  
 /note="assembly\_fragment"  
 14212. 16973  
 /note="assembly\_fragment"  
 17074. .19588  
 /note="assembly\_fragment"  
 19689. .22275  
 /note="assembly\_fragment"  
 22376. 25823  
 /note="assembly\_fragment"  
 25924. .31307  
 /note="assembly\_fragment"  
 31408. .34863  
 /note="assembly\_fragment"  
 34964. 41382  
 /note="assembly\_fragment"  
 41483. .49024  
 /note="assembly\_fragment"  
 49125. .58672  
 /note="assembly\_fragment"  
 58773. .69622  
 /note="assembly\_fragment"  
 69723. .88191  
 /note="assembly\_fragment"  
 88292. .107084  
 /note="assembly\_fragment"  
 107185. .153023  
 /note="assembly\_fragment"  
 clone\_end:T7  
 vector\_side:left  
 BASE COUNT 41494 a 35192 c 34573 g 39756 t 2008 others  
 ORIGIN

Query Match 2.4%; Score 20; DB 2; Length 153023;  
 Best Local Similarity 100.0%; Pred.No.19;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 614 TCAATCTAAATTTGCTTAC 633  
 |||||  
 Db 84374 TCAATCTAAATTTGCTTAC 84355

RESULT 37  
 AL353728/c 165739 bp DNA linear PRI 04-DEC-2001  
 LOCUS Human DNA sequence from clone RP11-402B2 on chromosome 9 Contains a

## ACCESSION

AL353728

## VERSION

AL353728.8 GI:11611340

## KEYWORDS

HTG.

## SOURCE

human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 165739)

## REFERENCE

Tracey,A.

## AUTHORS

Direct Submission

## TITLE

Submitted (04-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,

## JOURNAL

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

## COMMENT

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Dec 9, 2000 this sequence version replaced gi:11557901.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 only a small overlap as described above.

The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information  
 on the WORMPEP database can be found at

http://www.sanger.ac.uk/projects/c\_elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping  
 Group. Further information can be found at

http://www.sanger.ac.uk/MGP/Chr9  
 RP11-402B2 is from the library RPC1-11.2 constructed by the group  
 of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACe3.6

This sequence is the entire insert of clone RP11-402B2. The true  
 left end of clone RP11-572G24 is at 65138 in this sequence. This  
 sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest.

## FEATURES

## source

1. 165739  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="9"  
 /clone="RP11-402B2"  
 /clone\_lib="RPC1-11.2"  
 3. 418  
 /note="match: GSS: Em:AQ672631"

## misc\_feature

1097. 1457

## misc\_feature

/note="match: STS: Em:G53437

match: GSS: Em:AQ051700"

## repeat\_region

1599. 1664

## repeat\_region

/note="MLT1J repeat: matches 102. .167 of consensus"

## repeat\_region

/note="20 copies 2 mer ca 80% conserved"

## repeat\_region

3191. 3251

## repeat\_region

/note="MER41-internal repeat: matches 1095. .1155 of

## repeat\_region

consensus"

## repeat\_region

4046. 4325

## repeat\_region

/note="MLT1D repeat: matches 83. .334 of consensus"

## repeat\_region

6048. 6175

## repeat\_region

/note="MIR repeat: matches 21. .152 of consensus"

## repeat\_region

6252. 6480

## repeat\_region

/note="MIR repeat: matches 21. .254 of consensus"

## misc\_feature

7231. 7735

## repeat\_region

/note="match: GSS: Em:AQ568899"

## repeat\_region

7518. 7591

## repeat\_region

/note="MIR repeat: matches 65. .152 of consensus"

## repeat\_region

7670. 7713

```

/note="MLTIE repeat: matches 1. .44 of consensus"
7734. .7878
/note="MLTIE repeat: matches 45. .223 of consensus"
7796. .7911
/note="MLTIE repeat: matches 269. .379 of consensus"
7912. .8281
/note="MLTIB repeat: matches 1. .390 of consensus"
8282. .8627
/note="MLTIE repeat: matches 379. .568 of consensus"
11000. .11497
/note="match: GSS: Em:AQ751030"
11071. .11209
/note="11P10 repeat: matches 6012. .6159 of consensus"
complement(11173. .11405)
/note="match: GSS: Em:AQ060650"
complement(11197. .11531)
/note="match: GSS: Em:AQ308449"
11208. .11409
/note="match: GSS: Em:AQ474714 Em:AQ801225"
complement(11209. .11484)
/note="match: GSS: Em:AQ597874"
complement(11209. .11448)
/note="match: GSS: Em:AQ481695"
11210. .11622
/note="match: GSS: Em:AQ232833"
complement(11210. .11474)
/note="match: GSS: Em:B49798"
11210. .11253
/note="22 copies 2 mer at 86% conserved"
complement(11212. .11478)
/note="match: GSS: Em:AQ057475"
11213. .11471
/note="match: GSS: Em:AQ06817"
11213. .11322
/note="match: GSS: Em:AQ995542"
complement(11217. .11380)
/note="match: GSS: Em:AQ306702"
11219. .11319
/note="match: GSS: Em:AQ313645"
11244. .11423
/note="match: GSS: Em:AQ016479"
11282. .11502
/note="match: GSS: Em:AQ059847"
complement(11304. .11647)
/note="match: GSS: Em:AQ825390"
complement(11308. .11430)
/note="match: GSS: Em:AQ110479"
complement(11317. .11616)
/note="match: GSS: Em:AQ230662"
complement(11335. .11523)
/note="match: GSS: Em:AQ108237"
11344. .11537
/note="match: GSS: Em:AQ360911 Em:AQ362183"
complement(11349. .11690)
/note="match: GSS: Em:AQ294032"
complement(11383. .11493)
/note="match: GSS: Em:AQ101085"
complement(11375. .11516)
/note="match: STS: Em:DM100C58"
11377. .11639
/note="match: GSS: Em:B58206"
complement(11378. .11580)
/note="match: STS: Em:G12329"
complement(11390. .11657)
/note="match: GSS: Em:AQ986441"
complement(11393. .11696)
/note="match: GSS: Em:AQ811368"
11405. .11610
/note="match: GSS: Em:AQ234721"
complement(11417. .11536)
/note="match: GSS: Em:AQ932177"
complement(11429. .11790)
/note="match: GSS: Em:AQ401031"

misc_feature 11436. .11565
/note="match: GSS: Em:AQ412354"
11446. .11828
/note="match: GSS: Em:AQ473238"
11446. .11738
/note="match: GSS: Em:AQ424188"
complement(11446. .11686)
/note="match: GSS: Em:AQ472210"
11451. .11589
/note="match: GSS: Em:AQ421521"
11451. .11494
/note="22 copies 2 mer at 79% conserved"
complement(11452. .11657)
/note="match: STS: Em:G45547"
11457. .11746
/note="match: GSS: Em:AQ685385"
11457. .11650
/note="match: GSS: Em:AQ321208"
complement(11457. .11627)
/note="match: STS: Em:HS351WB1"
11460. .11879
/note="match: GSS: Em:AQ051188"
11478. .11697
/note="match: GSS: Em:BI7811"
complement(11489. .11723)
/note="match: GSS: Em:AQ391274"
complement(11497. .11790)
/note="match: GSS: Em:AQ292099"
11499. .11534
/note="18 copies 2 mer at 83% conserved"
complement(11507. .11766)
/note="match: GSS: Em:AQ334731"
11516. .11763
/note="match: GSS: Em:AQ513357"
complement(11521. .11628)
/note="match: GSS: Em:AQ133406"
complement(11523. .11674)
/note="match: GSS: Em:AQ317091"
complement(11575. .11724)
/note="match: GSS: Em:AQ477488"
complement(11580. .11765)
/note="match: GSS: Em:AQ264649"
complement(11582. .11815)
/note="match: GSS: Em:AQ346083"
complement(11585. .11722)
/note="match: GSS: Em:AQ460733"
complement(11601. .11800)
/note="match: GSS: Em:AQ578443"
complement(11602. .11799)
/note="match: GSS: Em:AQ259108"
11650. .11752
/note="match: GSS: Em:AQ572988"

Query Match 2.4%; Score 20; DB 9; Length 165739;
Best Local Similarity 100.0%; Pred.No.19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 643 ATAGGTATTAGTTACCTTAT 662
Db 6389 ATAGGTATTAGTTACCTTAT 6370

RESULT 38
AC048369/c
LOCUS AC048369 .166233 bp DNA linear HTG 25-JUN-2000
DEFINITION Homo sapiens chromosome 1 clone RP11-284G5 map 1, WORKING DRAFT
SEQUENCE, 29 unordered pieces.
ACCESSION AC048369
VERSION AC048369.2 GI:8705131
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```



REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 166233)  
Birren, B., Linton, L., Nussbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bada, F.,  
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,  
Campoliano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
Collins, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S.,  
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
McCarthy, M., McSwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
Meldrum, J., Meneus, D., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Tefaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (14-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 25, 2000 this sequence version replaced gi:7549699.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE  
JOURNAL  
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

Project Information

Center project name: L8045

Center clone name: 284\_G\_5

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 151910 bases at least Q40

Consensus quality: 158917 bases at least Q30

Consensus quality: 161704 bases at least Q20

Insert size: 178000; agarose-fp

Insert size: 163433; sum-of-contigs

Quality coverage: 3.7 in Q20 bases; agarose-fp

Quality coverage: 4.0 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently  
consists of 29 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 1069: contig of 1069 bp in length  
\* 1070 1169: gap of 100 bp  
\* 1170 2738: contig of 1569 bp in length  
\* 2739 2838: gap of 100 bp  
\* 2839 4143: contig of 1305 bp in length  
\* 4144 4243: gap of 100 bp  
\* 4244 5489: contig of 1246 bp in length  
\* 5490 5589: gap of 100 bp  
\* 5590 6873: contig of 1284 bp in length  
\* 6874 6973: gap of 100 bp  
\* 6974 9153: contig of 2180 bp in length

9154 9253: gap of 100 bp  
\* 9254 10555: contig of 1302 bp in length  
\* 10556 10655: gap of 100 bp  
\* 10656 13995: contig of 3340 bp in length  
\* 13996 14095: gap of 100 bp  
\* 14096 16689: contig of 2594 bp in length  
\* 16690 16789: gap of 100 bp  
\* 16790 19295: contig of 2506 bp in length  
\* 19296 19395: gap of 100 bp  
\* 19396 22797: contig of 3402 bp in length  
\* 22798 22897: gap of 100 bp  
\* 22898 26453: contig of 3556 bp in length  
\* 26454 26553: gap of 100 bp  
\* 26554 31369: contig of 4816 bp in length  
\* 31370 31469: gap of 100 bp  
\* 31470 35528: contig of 4059 bp in length  
\* 35529 35628: gap of 100 bp  
\* 35629 40531: contig of 4903 bp in length  
\* 40532 40631: gap of 100 bp  
\* 40632 45154: contig of 4523 bp in length  
\* 45155 45254: gap of 100 bp  
\* 45255 50400: contig of 5146 bp in length  
\* 50401 50500: gap of 100 bp  
\* 50501 56357: contig of 5857 bp in length  
\* 56358 56457: gap of 100 bp  
\* 56458 62667: contig of 6210 bp in length  
\* 62668 62767: gap of 100 bp  
\* 62768 69282: contig of 6515 bp in length  
\* 69283 69382: gap of 100 bp  
\* 69383 76270: contig of 8888 bp in length  
\* 76271 76370: gap of 100 bp  
\* 76371 84430: contig of 8060 bp in length  
\* 84431 84530: gap of 100 bp  
\* 84531 92060: contig of 7530 bp in length  
\* 92061 92180: gap of 100 bp  
\* 92181 101111: contig of 8951 bp in length  
\* 101112 101211: gap of 100 bp  
\* 101212 109009: contig of 7798 bp in length  
\* 109010 109109: gap of 100 bp  
\* 109110 120552: contig of 11443 bp in length  
\* 120553 120552: gap of 100 bp  
\* 120553 130699: contig of 10047 bp in length  
\* 130700 130799: gap of 100 bp  
\* 130800 145936: contig of 15137 bp in length  
\* 145937 146036: gap of 100 bp  
\* 146037 166233: contig of 20197 bp in length.

FEATURES  
Location/Qualifiers  
1. 166233  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1"  
/clone="RP11-284G5"  
/clone\_lib="RPC1-11 Human Male BAC"  
misc\_feature  
1. 1069  
/note="assembly\_fragment"  
misc\_feature  
1170. 2738  
/note="assembly\_fragment"  
misc\_feature  
2839. 4143  
/note="assembly\_fragment"  
misc\_feature  
4244. 5489  
/note="assembly\_fragment"  
misc\_feature  
5590. 6873  
/note="assembly\_fragment"  
misc\_feature  
6974. 9153  
/note="assembly\_fragment"  
misc\_feature  
9254. 10555  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right  
10856. 13995  
/note="assembly\_fragment"  
14096. 16689

```

misc_feature /note="assembly_fragment"
16790..19295
misc_feature /note="assembly_fragment"
19396..22797
misc_feature /note="assembly_fragment"
22898..26453
misc_feature /note="assembly_fragment"
26554..31369
misc_feature /note="assembly_fragment"
31470..35528
misc_feature /note="assembly_fragment"
35629..40531
misc_feature /note="assembly_fragment"
40632..45154
misc_feature /note="assembly_fragment"
45255..50400
misc_feature /note="assembly_fragment"
50501..56357
misc_feature /note="assembly_fragment"
56458..62667
misc_feature /note="assembly_fragment"
62768..69282
misc_feature /note="assembly_fragment"
69383..76270
misc_feature /note="assembly_fragment"
76371..84430
misc_feature /note="assembly_fragment"
84531..92060
misc_feature /note="assembly_fragment"
92161..101111
misc_feature /note="assembly_fragment"
101212..109009
misc_feature /note="assembly_fragment"
109110..120552
misc_feature /note="assembly_fragment"
120653..130699
misc_feature /note="assembly_fragment"
130800..145936
misc_feature /note="assembly_fragment"
146037..166233
misc_feature /note="assembly_fragment"
clone_end:SP6
vector:side:right"

BASE COUNT 44228 a 38868 c 38118 g 42215 t 2804 others
ORIGIN

Query Match 2.4%; Score 20; DB 2; Length 166233;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 614 TCAATCTAAATTTGCTTAC 633
DB 32277 TCAATCTAAATTTGCTTAC 32258

RESULT 39
AL606489/c
LOCUS Human DNA sequence from clone RP11-739N20 on chromosome 1, complete
sequence.
DEFINITION AL606489.25 GI:22204251
VERSION AL606489
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 167076)
AUTHORS Pearce,A.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Aug 11, 2002 this sequence version replaced gi:22002660.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
RP11-739N20 is from the library RPCI-11.3 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.
Location/Qualifiers
1..167076
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-739N20"
/clone_lib="RPCI-11.3"
BASE COUNT 46456 a 40657 c 37949 g 42014 t
ORIGIN

Query Match 2.4%; Score 20; DB 9; Length 167076;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 614 TCAATCTAAATTTGCTTAC 633
DB 155717 TCAATCTAAATTTGCTTAC 155698

RESULT 40
CNS06C7W/c
LOCUS Human chromosome 14 DNA sequence BAC R-572M18 of library RPCI-11
from chromosome 14 of Homo sapiens (Human), complete sequence.
DEFINITION CNS06C7W 177032 bp DNA linear PRI 29-APR-2001
ACCESSION AL390894
VERSION AL390894.2 GI:12001757
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 177032)
AUTHORS Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
Brottier,P., Catolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
Gyapay,G., Saurin,W. and Weissenbach,J.
TITLE Sequencing of the human chromosome 14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 177032)
AUTHORS Genoscope.

```

```

TITLE      Direct Submission
JOURNAL    Submitted (28-APR-2001) Genoscope - Centre National de Sequencage :
           BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT    - Web : www.genoscope.cns.fr)
           On Jan 1, 2001 this sequence version replaced gi:9650629.
           ----- Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr
           -----
The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-560013 (AC=AL391158)
Downstream BAC (overlapping the SP6 end) : R-630G18 -----
Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 7.17x in Q20 bases; sum-of-contigs
           -----
Overall quality chart :
Range      : bases
0          :
1 - 9      :
10 - 19    :
20 - 29    :
30 - 39    :      48
40 - 49    :      865
50 - 59    :      2434
60 - 69    :      7086
70 - 79    :      24396
80 - 89    :      61841
90 - 99    :      80362
           -----
Percentage of bases with a quality value >= 40 : 99 %.
FEATURES
source
1. 177032
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="14"
   /clone="R-572M18"
   /clone_lib="RPCI-11"
   34825_-35007
   /note="matching EMBL:H81577
   RHdb:RH44618
   dbSTS:STS37686
   Identified using the e-PCR software (G. Schuler)"
   34916_-35065
   /note="matching EMBL:H81577
   RHdb:RH53838
   RHdb:RH33524
   dbSTS:STS2795
   Identified using the e-PCR software (G. Schuler)"

BASE COUNT  57170 a 30035 c 30260 g 59567 t
ORIGIN

Query Match      2.4%; Score 20; DB 9; Length 177032;
Best Local Similarity 100.0%; Pred.No.19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      243 AACAAAAAAGACGATTTTA 262
Db      96306 AACAAAAAAGACGATTTTA 96287
           |||||
Search completed: July 4, 2003, 02:15:35
Job time : 1719 secs

```

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 01:03:12 ; Search time 188 Seconds  
(without alignments)  
10062.120 Million cell updates/sec

Title: US-10-062-624-41  
Perfect score: 840  
Sequence: 1 atgaatataagaataattct.....ttggaatgaggttcaccttc 840

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : N\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	840	100.0	840	24	ABK68876
2	371	44.2	495	20	AAK34770
3	30	3.6	852	20	AAK34744
4	22	2.6	22	24	ABK68873
5	19	2.3	206	21	AAC14373
6	19	2.3	840	21	AAD01294
7	19	2.3	840	24	ABK68854
8	19	2.3	843	20	AAK34762
9	19	2.3	894	20	AAK34749
-----					
1	840	100.0	840	24	ABK68876
2	371	44.2	495	20	AAK34770
3	30	3.6	852	20	AAK34744
4	22	2.6	22	24	ABK68873
5	19	2.3	206	21	AAC14373
6	19	2.3	840	21	AAD01294
7	19	2.3	840	24	ABK68854
8	19	2.3	843	20	AAK34762
9	19	2.3	894	20	AAK34749
-----					
1	840	100.0	840	24	ABK68876
2	371	44.2	495	20	AAK34770
3	30	3.6	852	20	AAK34744
4	22	2.6	22	24	ABK68873
5	19	2.3	206	21	AAC14373
6	19	2.3	840	21	AAD01294
7	19	2.3	840	24	ABK68854
8	19	2.3	843	20	AAK34762
9	19	2.3	894	20	AAK34749
-----					
1	840	100.0	840	24	ABK68876
2	371	44.2	495	20	AAK34770
3	30	3.6	852	20	AAK34744
4	22	2.6	22	24	ABK68873
5	19	2.3	206	21	AAC14373
6	19	2.3	840	21	AAD01294
7	19	2.3	840	24	ABK68854
8	19	2.3	843	20	AAK34762
9	19	2.3	894	20	AAK34749
-----					
1	840	100.0	840	24	ABK68876
2	371	44.2	495	20	AAK34770
3	30	3.6	852	20	AAK34744
4	22	2.6	22	24	ABK68873
5	19	2.3	206	21	AAC14373
6	19	2.3	840	21	AAD01294
7	19	2.3	840	24	ABK68854
8	19	2.3	843	20	AAK34762
9	19	2.3	894	20	AAK34749
-----					
1	840	100.0	840	24	ABK68876
2	371	44.2	495	20	AAK34770
3	30	3.6	852	20	AAK34744
4	22	2.6	22	24	ABK68873
5	19	2.3	206	21	AAC14373
6	19	2.3	840	21	AAD01294
7	19	2.3	840	24	ABK68854
8	19	2.3	843	20	AAK34762
9	19	2.3	894	20	AAK34749
-----					
1	840	100.0	840	24	ABK68876
2	371	44.2	495	20	AAK34770
3	30	3.6	852	20	AAK34744
4	22	2.6	22	24	ABK68873
5	19	2.3	206	21	AAC14373
6	19	2.3	840	21	AAD01294
7	19	2.3	840	24	ABK68854
8	19	2.3	843	20	AAK34762
9	19	2.3	894	20	AAK34749
-----					
1	840	100.0	840	24	ABK68876
2	371	44.2	495	20	AAK34770
3	30	3.6	852	20	AAK34744
4	22	2.6	22	24	ABK68873
5	19	2.3	206	21	AAC14373
6	19	2.3	840	21	AAD01294
7	19	2.3	840	24	ABK68854
8	19	2.3	843	20	AAK34762
9	19	2.3	894	20	AAK34749
-----					
1	840	100.0	840	24	ABK68876
2	371	44.2	495	20	AAK34770
3	30	3.6	852	20	AAK34744
4	22	2.6	22	24	ABK68873
5	19	2.3	206	21	AAC14373
6	19	2.3	840	21	AAD01294
7	19	2.3	840	24	ABK68854
8	19	2.3	843	20	AAK34762
9	19	2.3	894	20	AAK34749
-----					
1	840	100.0	840	24	ABK68876
2	371	44.2	495	20	AAK34770
3	30	3.6	852	20	AAK34744
4	22	2.6	22	24	ABK68873
5	19	2.3	206	21	AAC14373
6	19	2.3	840	21	AAD01294
7	19	2.3	840	24	ABK68854
8	19	2.3	843	20	AAK34762
9	19	2.3	894	20	AAK34749
-----					
1	840	100.0	840	24	ABK68876
2	371	44.2	495	20	AAK34770
3	30	3.6	852	20	AAK34744
4	22	2.6	22	24	ABK68873
5	19	2.3	206	21	AAC14373
6	19	2.3	840	21	AAD01294
7	19	2.3	840	24	ABK68854
8	19	2.3	843	20	AAK34762
9	19	2.3	894	20	AAK34749
-----					
1	840	100.0	840	24	ABK68876
2	371	44.2	495	20	AAK34770
3	30	3.6	852	20	AAK34744
4	22	2.6	22	24	ABK68873
5	19	2.3	206	21	AAC14373
6	19	2.3	840	21	AAD01294
7	19	2.3	840	24	ABK68854
8	19	2.3	843	20	AAK34762
9	19	2.3	894	20	AAK34749
-----					
1	840	100.0	840	24	ABK68876
2	371	44.2	495	20	AAK34770
3	30	3.6	852	20	AAK34744
4	22	2.6	22	24	ABK68873
5	19	2.3	206	21	AAC14373
6	19	2.3	840	21	AAD01294
7	19	2.3	840	24	ABK68854
8	19	2.3	843	20	AAK34762
9	19	2.3	894	20	AAK34749
-----					
1	840	100.0	840	24	ABK68876
2	371	44.2	495	20	AAK34770
3	30	3.6	852	20	AAK34744
4	22	2.6	22	24	ABK68873
5	19	2.3	206	21	AAC14373
6	19	2.3	840	21	AAD01294
7	19	2.3	840	24	ABK68854
8	19	2.3	843	20	AAK34762
9	19	2.3	894	20	AAK34749
-----					
1	840	100.0	840	24	ABK68876
2	371	44.2	495	20	AAK34770
3	30	3.6	852	20	AAK34744
4	22	2.6	22	24	ABK68873
5	19	2.3	206	21	AAC14373
6	19	2.3	840	21	AAD01294
7	19	2.3	840	24	ABK68854
8	19	2.3	843	20	AAK34762
9	19	2.3	894	20	AAK34749
-----					
1	840	100.0	840	24	ABK68876
2	371	44.2	495	20	AAK34770
3	30	3.6	852	20	AAK34744
4	22	2.6	22	24	ABK68873
5	19	2.3	206	21	AAC14373
6	19	2.3	840	21	AAD01294
7	19	2.3	840	24	ABK68854
8	19	2.3	843	20	AAK34762
9	19	2.3	894	20	AAK34749
-----					
1	840	100.0	840	24	ABK68876
2	371	44.2	495	20	AAK34770
3	30	3.6	852	20	AAK34744
4	22	2.6	22	24	ABK68873
5	19	2.3	206	21	AAC14373
6	19	2.3	840	21	AAD01294
7	19	2.3	840	24	ABK68854
8	19	2.3	843	20	AAK34762
9	19	2.3	894	20	AAK34749
-----					
1	840	100.0	840	24	ABK68876
2	371	44.2	495	20	AAK34770
3	30	3.6	852	20	AAK34744
4	22	2.6	22	24	ABK68873
5	19	2.3	206	21	AAC14373
6	19	2.3	840	21	AAD01294
7	19	2.3	840	24	ABK68854
8	19	2.3	843	20	AAK34762
9	19	2.3	894	20	AAK34749
-----					
1	840	100.0	840	24	ABK68876
2	371	44.2	495	20	AAK34770
3	30	3.6	852	20	AAK34744
4	22	2.6	22	24	ABK68873
5	19	2.3	206	21	AAC14373
6	19	2.3	840	21	AAD01294
7	19	2.3	840	24	ABK68854
8	19	2.3	843	20	AAK34762
9	19	2.3	894	20	AAK34749
-----					
1	840	100.0	840	24	ABK68876
2	371	44.2	495	20	AAK34770
3	30				

C 83	17	2.0	357	22	AAF31952	KGF-2 deletion con	C 156	17	2.0	555	20	AAZ11170	E.coli optimised K
C 84	17	2.0	402	20	AAZ11131	E.coli optimised	C 157	17	2.0	555	21	AAV71215	Human KGF-2 constr
C 85	17	2.0	402	21	AAV71249	Human KGF-2 delcti	C 158	17	2.0	555	22	AAF31913	POE60-Cys37 constr
C 86	17	2.0	402	22	AAF31947	KGF-2 deletion con	C 159	17	2.0	599	23	ABV56092	Human prostate exp
C 87	17	2.0	410	23	ABV44715	Human prostate exp	C 160	17	2.0	623	22	AAH00735	Enterococcus sulfu
C 88	17	2.0	411	23	ABV39927	Human prostate exp	C 161	17	2.0	627	19	AAV19977	E.coli optimised K
C 89	17	2.0	426	19	AAV20039	Mutated KGF-2 codi	C 162	17	2.0	627	20	AAZ11114	E.coli optimised
C 90	17	2.0	426	19	AAV20042	Mutated KGF-2 codi	C 163	17	2.0	627	21	AAV71223	Human KGF-2 E.coli
C 91	17	2.0	426	19	AAV20045	Mutated KGF-2 codi	C 164	17	2.0	627	22	AAF31921	Escherichia coli o
C 92	17	2.0	426	19	AAV20050	Mutated KGF-2 codi	C 165	17	2.0	663	24	ABK80374	Bacillus clausii g
C 93	17	2.0	426	19	AAV20024	Mutated KGF-2 codi	C 166	17	2.0	700	22	AAH92435	Human inflammatory
C 94	17	2.0	426	19	AAV20030	Mutated KGF-2 codi	C 167	17	2.0	762	24	ABK95237	Human prostate spe
C 95	17	2.0	426	19	AAV20033	Mutated KGF-2 codi	C 168	17	2.0	779	21	AAV59174	Human secreted pro
C 96	17	2.0	426	19	AAV20036	Mutated KGF-2 codi	C 169	17	2.0	779	22	ABA06231	Soy bean SCN/SCS r
C 97	17	2.0	426	19	AAV20023	Mutated KGF-2 codi	C 170	17	2.0	825	22	ABAL3545	Human nervous syst
C 98	17	2.0	426	19	AAZ11163	E.coli optimised K	C 171	17	2.0	828	24	AAD23932	Proteus mirabilis
C 99	17	2.0	426	20	AAZ11164	E.coli optimised K	C 172	17	2.0	867	20	AAV34759	DNA encoding P30 p
C 100	17	2.0	426	20	AAZ11165	E.coli optimised K	C 173	17	2.0	1130	24	AAI67641	F. necrophorum tru
C 101	17	2.0	426	20	AAZ11166	E.coli optimised K	C 174	17	2.0	1139	18	AAV79693	BRCA2 cancer susce
C 102	17	2.0	426	20	AAZ11167	E.coli optimised K	C 175	17	2.0	1164	18	AAV79692	BRCA2 cancer susce
C 103	17	2.0	426	20	AAZ11168	E.coli optimised K	C 176	17	2.0	1199	18	AAV79691	BRCA2 cancer susce
C 104	17	2.0	426	20	AAZ11162	E.coli optimised K	C 177	17	2.0	1607	21	AAD01292	Ehrlichia canis im
C 105	17	2.0	426	20	AAZ11096	KGF-2 deletion var	C 178	17	2.0	1607	24	ABK68852	DNA encoding Ehrli
C 106	17	2.0	426	20	AAZ11158	E.coli optimised K	C 179	17	2.0	1672	22	ABAL9785	Human nervous syst
C 107	17	2.0	426	20	AAZ11158	DNA encoding kerat	C 180	17	2.0	1672	22	AAZ30680	DNA encoding novel
C 108	17	2.0	426	20	AAZ11158	DNA encoding human	C 181	17	2.0	1672	22	AAZ28759	Genomic sequence #
C 109	17	2.0	426	21	AAV71268	Human KGF-2 mutant	C 182	17	2.0	1692	21	AAA69174	Bacteriophage Dp-1
C 110	17	2.0	426	21	AAV71283	Human KGF-2 mutant	C 183	17	2.0	1706	24	ABL49490	Sequence #92 used
C 111	17	2.0	426	21	AAV71304	Human KGF2delta33,	C 184	17	2.0	1706	24	ABK30680	Plant dwarfing/stu
C 112	17	2.0	426	21	AAV71292	Human KGF2delta33,	C 185	17	2.0	2005	22	AAV81805	Human secreted pro
C 113	17	2.0	426	21	AAV71295	Human KGF2delta33,	C 186	17	2.0	2044	18	AAV75623	Canine herpesvirus
C 114	17	2.0	426	21	AAV71298	Human KGF2delta33,	C 187	17	2.0	2044	19	AAV66945	Canine herpes viru
C 115	17	2.0	426	21	AAV71301	Human KGF2delta33,	C 188	17	2.0	2044	19	AAV66946	Canine herpes viru
C 116	17	2.0	426	21	AAV71304	Human KGF2delta33,	C 189	17	2.0	2044	22	AAF26765	Canine herpes viru
C 117	17	2.0	426	21	AAV71309	Human KGF2delta33,	C 190	17	2.0	2044	22	AAF26766	Canine herpes viru
C 118	17	2.0	426	22	AAF31966	KGF-2 mutant #3 co	C 191	17	2.0	2236	23	ABL05697	Drosophila melanog
C 119	17	2.0	426	22	AAF31981	KGF-2 mutant #4 co	C 192	17	2.0	2351	23	ABL15596	Drosophila melanog
C 120	17	2.0	426	22	AAF31987	KGF-2 mutant #6 co	C 193	17	2.0	2412	23	ABL26354	Drosophila melanog
C 121	17	2.0	426	22	AAF31990	KGF-2 mutant #8 co	C 194	17	2.0	2454	22	ABAI4614	Human nervous syst
C 122	17	2.0	426	22	AAF31993	KGF-2 mutant #9 co	C 195	17	2.0	2623	22	AADI7484	Mouse glucokinase
C 123	17	2.0	426	22	AAF31996	KGF-2 mutant #10 co	C 196	17	2.0	2664	15	AAQ56760	Chitinase derivati
C 124	17	2.0	426	22	AAF31999	KGF-2 mutant #11 c	C 197	17	2.0	2780	24	AAI67642	F. necrophorum tru
C 125	17	2.0	426	22	AAF32002	KGF-2 mutant #12 c	C 198	17	2.0	2989	18	AAV46203	Heliothis virescen
C 126	17	2.0	426	22	AAF32007	Mutant human KGF-2	C 199	17	2.0	3040	16	AAQ86783	DNA encoding Phosp
C 127	17	2.0	426	22	AAAC92945	Codon-optimised mu	C 200	17	2.0	3040	17	AAV42853	Phospholipase D pr
C 128	17	2.0	426	22	AAAC92960	E.coli optimised	C 201	17	2.0	3040	18	AAV85509	Phospholipase D en
C 129	17	2.0	444	20	AAZ11130	Human KGF-2 delcti	C 202	17	2.0	3047	18	AAV46204	Heliothis virescen
C 130	17	2.0	444	21	AAV71248	Escherichia coli o	C 203	17	2.0	3054	21	AAZ94068	Pneumocystis carin
C 131	17	2.0	447	22	AAF31946	Escherichia coli o	C 204	17	2.0	3309	23	ABL17564	Drosophila melanog
C 132	17	2.0	447	22	AAF32037	Mutated KGF-2 codi	C 205	17	2.0	4373	23	ABL17542	Drosophila melanog
C 133	17	2.0	456	22	AAF32034	Mutated KGF-2 codi	C 206	17	2.0	4897	23	ABL05696	Drosophila melanog
C 134	17	2.0	516	19	AAV20027	E.coli optimised K	C 207	17	2.0	5504	24	ABL70572	Chemically treated
C 135	17	2.0	516	20	AAZ11161	Human KGF-2 mutant	C 208	17	2.0	5504	24	AAV61256	Human gene regulat
C 136	17	2.0	516	21	AAV71286	KGF-2 mutant #5 co	C 209	17	2.0	5504	24	ABK31351	Signal transductio
C 137	17	2.0	516	22	AAF31984	E.coli optimised K	C 210	17	2.0	5534	12	AAQ13574	EGR gene. Drosoph
C 138	17	2.0	525	19	AAV19978	E.coli optimised	C 211	17	2.0	5895	18	AAV74312	Staphylococcus aur
C 139	17	2.0	525	20	AAZ11138	E.coli optimised	C 212	17	2.0	6396	24	ABL34247	Human immune syste
C 140	17	2.0	525	20	AAZ11139	E.coli optimised	C 213	17	2.0	6654	23	ABLO3460	Drosophila melanog
C 141	17	2.0	525	20	AAZ11129	Truncated E.coli	C 214	17	2.0	6756	23	ABL30434	Drosophila melanog
C 142	17	2.0	525	20	AAZ11117	Human KGF-2 E.coli	C 215	17	2.0	8067	24	ABN80225	Human chemically m
C 143	17	2.0	525	21	AAV71226	Human KGF-2 delcti	C 216	17	2.0	8132	23	ABL09598	Drosophila melanog
C 144	17	2.0	525	21	AAV71247	Human KGF-2 cystei	C 217	17	2.0	9500	24	AAD23931	Proteus mirabilis
C 145	17	2.0	525	21	AAV71256	Human KGF-2 cystei	C 218	17	2.0	9726	24	AAI67640	F. necrophorum leu
C 146	17	2.0	525	21	AAV71257	Escherichia coli o	C 219	17	2.0	9903	22	AAAL05500	Human reproductive
C 147	17	2.0	525	22	AAF31924	KGF-2 deletion con	C 220	17	2.0	10891	24	ABL32465	Human immune syste
C 148	17	2.0	525	22	AAF31945	KGF-2 mutant const	C 221	17	2.0	11130	24	AAI67647	F. necrophorum leu
C 149	17	2.0	525	22	AAF31954	KGF-2 mutant const	C 222	17	2.0	11474	23	ABK42672	Genomic sequence #
C 150	17	2.0	525	22	AAF31955	Oligonucleotide fo	C 223	17	2.0	13158	18	AAV75288	Nucleotide sequenc
C 151	17	2.0	531	24	ABQ22992	Oligonucleotide fo	C 224	17	2.0	14117	24	ABL64107	Breast cancer rela
C 152	17	2.0	531	24	ABQ22993	Human prostate exp	C 225	17	2.0	14552	23	ABLO3818	Drosophila melanog
C 153	17	2.0	536	23	ABV01367	Bacillus lichenifo	C 226	17	2.0	15413	22	AAK84002	Human immune/haema
C 154	17	2.0	543	24	ABK73249	Human cancer relat	C 227	17	2.0	17450	22	AAAL05953	Human reproductive
C 155	17	2.0	545	24	ABN62806		C 228	17	2.0	17450	22	AAK70834	Human immune/haema

229	17	2.0	17450	23	ABL98517	Human testicular a	302	16	1.9	587	22	AAK34623	Human bone marrow
c 230	17	2.0	17959	24	ABL54342	Chemically treated	303	16	1.9	587	22	AAI16795	Probe #6728 for ge
c 231	17	2.0	17959	24	ABL32575	Human immune syste	304	16	1.9	587	22	AAI140342	Probe #9028 used t
c 232	17	2.0	18855	22	ABL32610	Human immune syste	305	16	1.9	587	24	ABO19245	Human genome-deriv
c 233	17	2.0	23419	22	AAS35864	Human cardiovascular	c 306	16	1.9	587	24	ABO19245	Mouse ischaemic co
c 234	17	2.0	26110	22	AAK65036	Human immune/haema	c 307	16	1.9	595	23	ABV45668	Human prostate exp
c 235	17	2.0	26110	22	AAK78526	Human immune/haema	c 308	16	1.9	604	24	ABL49445	Sequence #47 used
c 236	17	2.0	29871	24	ABN86359	L. lactis multidru	c 309	16	1.9	604	24	ABK30632	Plant dwarfing/stu
c 237	17	2.0	44400	22	AAS12438	DNA encoding 1-ami	c 310	16	1.9	604	24	ABK30882	Plant dwarfing/stu
c 238	17	2.0	48551	24	AAS20800	Clostridium diffic	c 311	16	1.9	616	24	ABK30797	Plant dwarfing/stu
c 239	17	2.0	56506	21	AAAG61968	Bacteriophage Dp-1	c 312	16	1.9	616	24	ABK30984	Plant dwarfing/stu
c 240	17	2.0	335913	22	AAI61371	Soybean 240017 reg	c 313	16	1.9	627	20	AAK399613	Nucleic acid seque
c 241	17	2.0	335913	22	AAI61372	Soybean 240017 reg	c 314	16	1.9	633	24	ABN67778	Streptococcus poly
c 242	17	2.0	34980	22	AAH68532	C glutamicum codin	c 315	16	1.9	643	24	ABK33306	DNA encoding novel
c 243	17	2.0	465237	24	ABQ87681	Human oestrogen re	c 316	16	1.9	662	24	ABQ40374	Oligonucleotide fo
c 244	17	2.0	465237	24	ABA90193	Human oestrogen re	c 317	16	1.9	662	24	ABQ40375	Oligonucleotide fo
c 245	17	2.0	495269	24	ABQ67195	Listeria innocua c	c 318	16	1.9	670	21	AAQ67660	Human secreted pro
c 246	17	2.0	640681	24	ABA92787	Buchnera sp. genom	c 319	16	1.9	694	24	ABQ30448	Oligonucleotide fo
c 247	17	2.0	1038602	20	AAZ01425	Complete genome se	c 320	16	1.9	694	24	ABQ30449	Oligonucleotide fo
c 248	17	2.0	1082138	21	AAE22305	Arabidopsis thalia	c 321	16	1.9	697	24	ABQ43312	Oligonucleotide fo
c 249	17	2.0	1830121	17	AA742063	Haemophilus influe	c 322	16	1.9	697	24	ABQ43313	Oligonucleotide fo
c 250	17	2.0	2365589	24	ABA90521	Genomic sequence o	c 323	16	1.9	737	21	AAQ98892	Human pancreatic c
c 251	17	2.0	3011208	24	ABQ69245	Listeria innocua D	c 324	16	1.9	741	19	AAV59520	Human secreted pro
c 252	16	1.9	20	24	AAS98610	Ehrlichia chaffeen	c 325	16	1.9	742	20	AAK25137	Mouse cell cycle c
c 253	16	1.9	32	19	AAV07959	Helicobacter pylor	c 326	16	1.9	776	22	AAH06280	Human cDNA clone (
c 254	16	1.9	32	19	AAV07955	Helicobacter pylor	c 327	16	1.9	794	18	AAV75146	Staphylococcus aur
c 255	16	1.9	60	24	ABN44764	Human spliced tran	c 328	16	1.9	798	21	AAQ97334	Helicobacter pylor
c 256	16	1.9	155	16	AAI21466	Human gene signatu	c 329	16	1.9	802	22	AAH06608	Human cDNA clone (
c 257	16	1.9	175	20	AAK40985	Human secreted pro	c 330	16	1.9	816	23	AAQ66258	DNA encoding novel
c 258	16	1.9	186	24	ABN71417	Streptococcus poly	c 331	16	1.9	829	24	ABQ43870	Oligonucleotide fo
c 259	16	1.9	192	21	AAK02919	Human secreted pro	c 332	16	1.9	829	24	ABQ43871	Oligonucleotide fo
c 260	16	1.9	208	22	AAK44445	Human bone marrow	c 333	16	1.9	838	22	AAH04072	Human cDNA clone (
c 261	16	1.9	208	24	ABN18680	Human genome-deriv	c 334	16	1.9	849	24	ABN68167	Streptococcus poly
c 262	16	1.9	309	24	ABN25402	Human ORFX polynuc	c 335	16	1.9	900	12	AAQ11942	Proximal 5'DNA and
c 263	16	1.9	318	24	ABL67320	Thyroid cancer rel	c 336	16	1.9	900	13	AAQ29373	Mouse beta-RAR pro
c 264	16	1.9	318	24	ABL67320	Kidney cancer rela	c 337	16	1.9	911	21	AAQ33639	Zea mays DNA fragm
c 265	16	1.9	333	22	AAK37567	Novel human diagno	c 338	16	1.9	919	21	AAQ35670	Arabidopsis thalia
c 266	16	1.9	348	17	AAK30560	H. pylori cellular	c 339	16	1.9	937	24	ABK92408	Human prostate spe
c 267	16	1.9	357	17	AAI16699	Lewis-Y antibody B	c 340	16	1.9	981	19	AAK14061	H. pylori GPO 357
c 268	16	1.9	357	17	AAI16697	Lewis-Y antibody B	c 341	16	1.9	1000	21	AAH51496	Human UGT2B4 relat
c 269	16	1.9	357	17	AAI16696	Lewis-Y antibody B	c 342	16	1.9	1031	24	ABQ41106	Oligonucleotide fo
c 270	16	1.9	372	23	AAS65729	DNA encoding novel	c 343	16	1.9	1031	24	ABQ41107	Oligonucleotide fo
c 271	16	1.9	382	21	AAK94019	Cat flea hindgut a	c 344	16	1.9	1032	24	ABQ31732	Oligonucleotide fo
c 272	16	1.9	385	23	ABV04532	Human prostate exp	c 345	16	1.9	1032	24	ABQ31733	Oligonucleotide fo
c 273	16	1.9	393	23	AAS93402	DNA encoding novel	c 346	16	1.9	1064	22	AAI37497	Human musculoskele
c 274	16	1.9	423	22	AAI35595	Human musculoskele	c 347	16	1.9	1067	24	ABQ68549	Listeria monocytog
c 275	16	1.9	423	24	ABN77193	Human ORF2140 cDNA	c 348	16	1.9	1134	22	AAH84630	E. coli growth and
c 276	16	1.9	427	20	AAK51788	Human secreted pro	c 349	16	1.9	1141	18	AAV74837	Staphylococcus aur
c 277	16	1.9	431	21	AAK42151	Human secreted exp	c 350	16	1.9	1149	20	AAK99651	Nucleic acid seque
c 278	16	1.9	431	22	AAH33822	Human colon cancer	c 351	16	1.9	1170	19	AAK30561	H. pylori cellular
c 279	16	1.9	435	22	AAS30998	Human diagnostic a	c 352	16	1.9	1224	19	AAK14573	H. pylori GPO 162
c 280	16	1.9	441	21	AAK28433	Human secreted pro	c 353	16	1.9	1257	18	AAK83803	DNA encoding 2 Sta
c 281	16	1.9	442	21	AAS33457	Arabidopsis thalia	c 354	16	1.9	1297	21	AAK42171	Arabidopsis thalia
c 282	16	1.9	451	21	AAK41717	Human secreted exp	c 355	16	1.9	1299	23	AAK91717	DNA encoding novel
c 283	16	1.9	455	21	AAK22480	Human secreted pro	c 356	16	1.9	1312	21	AAK59439	Nucleotide sequenc
c 284	16	1.9	455	24	ABN20926	Human ORFX polynuc	c 357	16	1.9	1312	24	ABQ71185	Listeria monocytog
c 285	16	1.9	461	22	AAK56703	Human immune/haema	c 358	16	1.9	1332	21	AAK46192	Arabidopsis thalia
c 286	16	1.9	466	21	AAK94840	Cat flea hindgut a	c 359	16	1.9	1332	22	AAH53102	S. epidermidis ope
c 287	16	1.9	483	22	AAK31359	Human bone marrow	c 360	16	1.9	1332	24	ABN92615	Staphylococcus epi
c 288	16	1.9	483	23	AAS5049	DNA encoding novel	c 361	16	1.9	1335	21	AAK42036	Arabidopsis thalia
c 289	16	1.9	483	24	ABN06111	Human genome-deriv	c 362	16	1.9	1347	21	AAK75617	Nucleotide sequenc
c 290	16	1.9	501	24	ABN67276	Streptococcus poly	c 363	16	1.9	1347	24	AAK20732	Human zalphall Lig
c 291	16	1.9	502	18	AAI79663	BRCA2 cancer succe	c 364	16	1.9	1375	21	AAK51920	Arabidopsis thalia
c 292	16	1.9	523	22	AAH13571	Human cDNA clone (	c 365	16	1.9	1379	22	AAK84307	Human EXCS encodin
c 293	16	1.9	525	24	ABN63616	Human cancer relat	c 366	16	1.9	1380	21	AAK70124	Plasmodium falcipa
c 294	16	1.9	542	22	AAH28989	Drosophila melanog	c 367	16	1.9	1409	22	ABAI6382	Human nervous syst
c 295	16	1.9	562	23	ABV55553	Human prostate exp	c 368	16	1.9	1421	21	AAK39735	Potato cv. Cara in
c 296	16	1.9	563	22	ABAI1877	Human nervous syst	c 369	16	1.9	1429	22	AAH29726	S. cerevisiae apopt
c 297	16	1.9	573	24	ABN67339	Streptococcus poly	c 370	16	1.9	1453	21	AAFI3624	Aspergillus oryzae
c 298	16	1.9	586	23	ABV45655	Human prostate exp	c 371	16	1.9	1463	21	AAK46528	cDNA sequence enco
c 299	16	1.9	587	22	ABA60457	Human foetal liver	c 372	16	1.9	1463	24	ABK72650	DNA encoding human
c 300	16	1.9	587	22	ABA86659	Probe #125 for ge	c 373	16	1.9	1477	24	ABK72652	DNA encoding human
c 301	16	1.9	587	22	AAK08736	Human brain expres	c 374	16	1.9	1487	24	AAK99201	cDNA of Human - ho

C 375	16	1.9	1489	22	AAS27372	cDNA encoding nove	C 448	16	1.9	2642	23	ABL10041	Drosophila melanog
C 376	16	1.9	1489	22	AAS35017	cDNA encoding nove	C 449	16	1.9	2679	19	AAV35427	Brassica napus AGL
C 377	16	1.9	1512	23	ABL22001	Drosophila melanog	C 450	16	1.9	2699	19	AAV07912	Helicobacter pylor
C 378	16	1.9	1517	21	AAC37923	Arabidopsis thalia	C 451	16	1.9	2701	19	AAV31998	Flax SAD1 gene. L
C 379	16	1.9	1519	21	AAA26700	Candida albicans p	C 452	16	1.9	2705	19	AAV31999	Flax SAD2 gene. L
C 380	16	1.9	1534	21	AAA46531	cDNA sequence enco	C 453	16	1.9	2729	24	ABK35391	Human cDNA encodin
C 381	16	1.9	1534	24	ABK72653	DNA encoding rat p	C 454	16	1.9	2821	23	ABL23502	Drosophila melanog
C 382	16	1.9	1540	21	AAA46532	cDNA sequence enco	C 455	16	1.9	2864	22	AAV84499	Human aminopeptida
C 383	16	1.9	1540	24	ABK72654	DNA encoding mouse	C 456	16	1.9	2915	19	AAV07913	Helicobacter pylor
C 384	16	1.9	1564	24	AAS62626	cDNA sequence #413	C 457	16	1.9	3019	23	ABL09002	Drosophila melanog
C 385	16	1.9	1614	22	AAH98192	Human EST-derived	C 458	16	1.9	3024	24	ABN66444	Streptococcus poly
C 386	16	1.9	1617	22	AAL37498	Human musculoskele	C 459	16	1.9	3045	22	AAV05197	Human secreted pro
C 387	16	1.9	1617	23	ABK43047	Genomic sequence #	C 460	16	1.9	3115	19	AAV41374	RANK partial polyp
C 388	16	1.9	1639	24	ABK72693	DNA encoding human	C 461	16	1.9	3115	19	AAV41368	RANK partial polyp
C 389	16	1.9	1658	22	AAH14379	Human cDNA sequenc	C 462	16	1.9	3115	22	AAV15307	Human receptor act
C 390	16	1.9	1669	23	ABL07955	Drosophila melanog	C 463	16	1.9	3115	22	AAV08711	Human receptor act
C 391	16	1.9	1741	22	AAH17198	Human cDNA sequenc	C 464	16	1.9	3115	22	AAV05900	Human RANK (recept
C 392	16	1.9	1767	11	AAQ06900	Neo-pullulanase ge	C 465	16	1.9	3136	19	AAV41376	NF-kB receptor act
C 393	16	1.9	1768	20	AAH87686	Human tyrosylprote	C 466	16	1.9	3136	19	AAV41370	NF-kB receptor act
C 394	16	1.9	1768	21	AAZ99697	cDNA encoding a hu	C 467	16	1.9	3136	22	AAV15309	Human receptor act
C 395	16	1.9	1800	24	ABQ90220	M. capsulatus gene	C 468	16	1.9	3136	22	AAV08713	Human receptor act
C 396	16	1.9	1816	22	AAI60832	Human polynucleoti	C 469	16	1.9	3136	22	AAV05902	Human full-length
C 397	16	1.9	1831	22	AAI59046	Human polynucleoti	C 470	16	1.9	3136	22	AAV31745	Human RANK nucleot
C 398	16	1.9	1835	21	AAAC63914	Cucumber xylen sap	C 471	16	1.9	3136	24	AAK12878	Human TRANCE recep
C 399	16	1.9	1856	21	AAAC6529	cDNA sequence enco	C 472	16	1.9	3137	21	AAK236257	cDNA encoding a hu
C 400	16	1.9	1856	24	ABK72651	DNA encoding rat p	C 473	16	1.9	3154	22	AAV05906	Human FEO RANK (re
C 401	16	1.9	1856	24	ABK72704	DNA encoding rat p	C 474	16	1.9	3206	22	AAV05907	Human EP RANK (rec
C 402	16	1.9	1864	23	ABL18215	Drosophila melanog	C 475	16	1.9	3206	21	AAV46309	cDNA encoding an l
C 403	16	1.9	1867	20	AAH87687	Mouse tyrosylprote	C 476	16	1.9	3340	19	AAV04154	Blood group antige
C 404	16	1.9	1867	21	AAZ99698	cDNA encoding a mu	C 477	16	1.9	3364	23	ABL23238	Drosophila melanog
C 405	16	1.9	1907	21	AAAC6530	cDNA sequence enco	C 478	16	1.9	3406	21	AAV23439	cDNA encoding huma
C 406	16	1.9	1907	24	ABK72652	DNA encoding mouse	C 479	16	1.9	3448	23	ABL11425	Drosophila melanog
C 407	16	1.9	1916	24	ABK35788	cDNA sequence #179	C 480	16	1.9	3486	24	ABK12887	Human protease PRT
C 408	16	1.9	1919	22	AAK05010	Human brain expres	C 481	16	1.9	3539	20	AAZ24245	W09916900 Seq ID 1
C 409	16	1.9	1919	24	ABK35788	Human genome-deriv	C 482	16	1.9	3568	21	AAZ29491	DNA encoding a mai
C 410	16	1.9	1921	18	AAH58786	cDNA encoding C-CA	C 483	16	1.9	3581	22	AAV09377	Human vesicle traf
C 411	16	1.9	1923	21	AAAC4722	Arabidopsis thalia	C 484	16	1.9	3597	21	AAV42815	Arabidopsis thalia
C 412	16	1.9	1938	22	ABK16016	Human nervous syst	C 485	16	1.9	3600	20	AAK82940	Human immune/haema
C 413	16	1.9	1939	24	ABK70543	Listeria monocytog	C 486	16	1.9	3620	22	AAK15296	Sequence from plas
C 414	16	1.9	1944	22	ABK16023	Human nervous syst	C 487	16	1.9	3625	22	AAV05142	Human secreted pro
C 415	16	1.9	1949	22	ABK16003	Human nervous syst	C 488	16	1.9	3643	22	AAH54647	S. epidermidis gen
C 416	16	1.9	1949	22	ABK16381	Human nervous syst	C 489	16	1.9	3661	23	ABL22000	Drosophila melanog
C 417	16	1.9	1962	22	AAK03851	Marine cell cycle	C 490	16	1.9	3715	23	ABV30160	Human prostate exp
C 418	16	1.9	1962	24	AAH18422	cDNA encoding muri	C 491	16	1.9	3715	23	ABV30171	Human prostate exp
C 419	16	1.9	1975	22	AAV26736	Human breast cance	C 492	16	1.9	3738	23	ABL21954	Drosophila melanog
C 420	16	1.9	2035	22	AAV06049	Human reproductive	C 493	16	1.9	3760	20	AAV20279	Borrelia burgdorfe
C 421	16	1.9	2035	22	ABL98614	Human testicular a	C 494	16	1.9	3776	21	AAZ58263	Corn cellulose syn
C 422	16	1.9	2046	24	ABK56796	Nucleotide sequenc	C 495	16	1.9	3844	22	ABL02568	Drosophila melanog
C 423	16	1.9	2061	24	ABA04600	MOL9b coding sequ	C 496	16	1.9	3873	22	ABK19128	Human nervous syst
C 424	16	1.9	2093	21	AAV57996	2093 bp Candida al	C 497	16	1.9	3943	22	AAV31317	Human cDNA encodin
C 425	16	1.9	2115	24	ABK03922	Human NHP protein	C 498	16	1.9	3943	24	ABK56641	Human polynucleoti
C 426	16	1.9	2148	13	AAQ273707	Rabbit endopeptida	C 499	16	1.9	3945	22	AAV31428	DNA encoding novel
C 427	16	1.9	2167	24	ABA04601	MOL9c coding sequ	C 500	16	1.9	3945	23	ABK43764	DNA encoding novel
C 428	16	1.9	2170	24	ABK92040	DNA encoding novel	C 501	16	1.9	3945	24	ABK66752	Human polynucleoti
C 429	16	1.9	2170	24	ABK92041	DNA encoding novel	C 502	16	1.9	3975	21	AAV48384	Arabidopsis thalia
C 430	16	1.9	2170	24	ABK92042	DNA encoding novel	C 503	16	1.9	4116	24	ABK24522	BIF-2alpha kinase
C 431	16	1.9	2170	24	ABK92043	DNA encoding novel	C 504	16	1.9	4169	23	ABL16690	Drosophila melanog
C 432	16	1.9	2170	24	ABK92044	DNA encoding novel	C 505	16	1.9	4272	9	AAV80447	Modified factor VI
C 433	16	1.9	2170	24	ABK92045	DNA encoding novel	C 506	16	1.9	4275	9	AAV80446	Modified factor VI
C 434	16	1.9	2170	24	ABK92047	DNA encoding novel	C 507	16	1.9	4275	10	AAV30654	DNA encoding 740 A
C 435	16	1.9	2170	24	ABK92048	DNA encoding novel	C 508	16	1.9	4330	23	ABL11730	Drosophila melanog
C 436	16	1.9	2171	24	ABK92046	DNA encoding novel	C 509	16	1.9	4334	19	AAV12142	Homo sapiens facto
C 437	16	1.9	2239	21	AAV47186	Arabidopsis thalia	C 510	16	1.9	4334	20	AAV31195	Porcine factor VII
C 438	16	1.9	2307	24	ABN79890	Fungal ZBC gene se	C 511	16	1.9	4334	22	AAV05051	cDNA encoding porc
C 439	16	1.9	2355	24	ABA04599	MOL9a coding sequ	C 512	16	1.9	4372	23	ABL07954	Drosophila melanog
C 440	16	1.9	2367	23	ABL22265	Drosophila melanog	C 513	16	1.9	4373	20	AAV82258	Beta-domain delete
C 441	16	1.9	2371	22	AAH16341	Human cDNA sequenc	C 514	16	1.9	4404	22	AAV18667	Modified porcine f
C 442	16	1.9	2379	22	AAH16796	Human cDNA sequenc	C 515	16	1.9	4451	22	AAK68837	Human immune/haema
C 443	16	1.9	2430	21	AAV37972	Arabidopsis thalia	C 516	16	1.9	4545	9	AAV80444	Modified factor VI
C 444	16	1.9	2442	24	ABN68312	Streptococcus poly	C 517	16	1.9	4616	9	AAV81545	Human Factor VIII-
C 445	16	1.9	2466	11	AAQ06903	Sequence encoding	C 518	16	1.9	4629	16	AAQ76016	B-domain deleted F
C 446	16	1.9	2502	23	AAV81977	DNA encoding novel	C 519	16	1.9	4629	20	AAV88293	Human Factor VIII
C 447	16	1.9	2601	20	AAZ25013	Lactobacillus reut	C 520	16	1.9	4629	24	ABN84384	Human B domain-del





c 667	16	1.9	16602	24	ABN00069	Human chemically m	740	16	1.9	2155561	24	ABN71527	Streptococcus poly
c 668	16	1.9	16602	24	ABL32727	Human immune syste	c 741	16	1.9	2155561	24	ABN71527	Streptococcus poly
c 669	16	1.9	16811	24	ABL33946	Human immune syste	c 742	16	1.9	2365589	24	ABA90521	Genomic sequence o
c 670	16	1.9	17425	20	AX28366	Human Stat6 gene.	c 743	16	1.9	4403765	22	AAI99683	Mycobacterium tube
c 671	16	1.9	17495	22	AAK82938	Human immune/haema	c 744	16	1.9	4411529	22	AAI99682	Mycobacterium tube
c 672	16	1.9	17495	22	AAK82938	Human immune/haema	c 745	15	1.8	19	21	AAZ76207	Human biatlalic ma
c 673	16	1.9	17569	24	ABK39756	CNA encoding clon	c 746	15	1.8	19	21	AAH61111	Cdc 25 hs ribozyme
c 674	16	1.9	17594	24	ABL34026	Human immune syste	c 747	15	1.8	19	22	AAH61273	Cdc25 hs ribozyme
c 675	16	1.9	17594	24	ABL34027	Human immune syste	c 748	15	1.8	24	24	ABL40414	200 basepair nomol
c 676	16	1.9	17666	22	AAK76302	Human immune/haema	c 749	15	1.8	40	21	AAH08615	Primer 2 used to c
c 677	16	1.9	17893	24	ABL33364	Human immune/haema	c 750	15	1.8	47	19	AAV20007	PCR primer for mut
c 678	16	1.9	18011	24	ABL32034	Human immune syste	c 751	15	1.8	47	20	AAZ11125	PCR primer for E.c
c 679	16	1.9	18357	24	ABQ67084	Human angiogenesis	c 752	15	1.8	47	21	AAH71234	Human KGF-2 E. col
c 680	16	1.9	18585	24	ABL34609	Human metacasis a	c 753	15	1.8	47	22	AAH31932	PCR primer #10 for
c 681	16	1.9	18624	24	ABL33702	Human immune syste	c 754	15	1.8	48	19	AAV20003	PCR primer for mut
c 682	16	1.9	19301	21	AAH21245	Human low adenosin	c 755	15	1.8	48	20	AAZ11121	PCR primer for E.c
c 683	16	1.9	19301	21	AAA35123	Human adenosine re	c 756	15	1.8	48	21	AAH71230	Human KGF-2 E. col
c 684	16	1.9	19659	24	ABL33766	Human immune syste	c 757	15	1.8	48	22	AAH31928	PCR primer #6 for
c 685	16	1.9	21537	24	ABL33999	Human immune syste	c 758	15	1.8	51	24	ABA05855	Beta 3 endonexin c
c 686	16	1.9	21729	23	ABL06154	Drosophila melanog	c 759	15	1.8	60	24	ABN33328	Human spliced tran
c 687	16	1.9	22407	22	AAF28529	Genomic fragment #	c 760	15	1.8	70	17	AAH97729	Nucleic acid inhib
c 688	16	1.9	25461	23	ABL28860	Drosophila melanog	c 761	15	1.8	70	21	AAZ52638	High salt SLEEX hu
c 689	16	1.9	26241	22	ABA16222	Human nervous syst	c 762	15	1.8	103	24	ABN61389	Human cancer relat
c 690	16	1.9	27423	23	ABL04268	Drosophila melanog	c 763	15	1.8	107	21	AAH72823	Cosmid ZK643 repea
c 691	16	1.9	32134	22	AAI98172	Human excretory re	c 764	15	1.8	145	24	ABL76968	S solfataricus rpo
c 692	16	1.9	32134	22	AAI63522	Human kidney relat	c 765	15	1.8	147	24	ABQ69368	Listeria innocua D
c 693	16	1.9	32192	22	AAI99173	Human excretory re	c 766	15	1.8	149	22	ABA71141	Human foetal liver
c 694	16	1.9	32192	22	AAI63523	Human kidney relat	c 767	15	1.8	149	22	AAH19431	Human brain expres
c 695	16	1.9	33632	23	ABL28040	Drosophila melanog	c 768	15	1.8	149	22	AAH45418	Human bone marrow
c 696	16	1.9	34361	23	ABL18288	Drosophila melanog	c 769	15	1.8	149	22	AAI51364	Probe #20050 used
c 697	16	1.9	34712	23	ABL18214	Drosophila melanog	c 770	15	1.8	149	24	ABN19682	Human genome-deriv
c 698	16	1.9	35100	22	AAK65700	Human immune/haema	c 771	15	1.8	153	21	AAA43076	Human secreted exp
c 699	16	1.9	35100	22	AAK69767	Human immune/haema	c 772	15	1.8	158	16	AAH23645	Human gene signatu
c 700	16	1.9	35115	22	AAK65699	Human immune/haema	c 773	15	1.8	169	21	AAH31660	Human secreted pro
c 701	16	1.9	35115	22	AAK65699	Human immune/haema	c 774	15	1.8	171	21	AAA42248	Human secreted exp
c 702	16	1.9	35664	23	ABL03070	Drosophila melanog	c 775	15	1.8	174	24	ABN79359	Human ORF4306 cDNA
c 703	16	1.9	42519	23	ABK81318	Human immune/haema	c 776	15	1.8	183	16	AAH24680	Human gene signatu
c 704	16	1.9	44100	21	ABN97975	Human retroviral s	c 777	15	1.8	183	24	ABN33087	Staphylococcus epi
c 705	16	1.9	44242	23	ABL19930	Drosophila melanog	c 778	15	1.8	184	21	AAH25288	Human secreted pro
c 706	16	1.9	46366	22	AAK82098	Human immune/haema	c 779	15	1.8	187	21	AAA41529	Human secreted exp
c 707	16	1.9	48551	24	AAH20800	Clostridium diffi	c 780	15	1.8	189	21	AAA44745	Human secreted exp
c 708	16	1.9	56093	24	ABL61744	Colon adenocarcino	c 781	15	1.8	194	24	ABQ73039	Human chloride cha
c 709	16	1.9	61890	23	ABL16586	Drosophila melanog	c 782	15	1.8	201	22	ABA71762	Human foetal liver
c 710	16	1.9	72215	22	AAK86832	Human immune/haema	c 783	15	1.8	201	22	ABA37848	Probe #16314 for g
c 711	16	1.9	72928	20	AAZ18355	Human ASTHJ 5' ge	c 784	15	1.8	201	22	AAK20128	Human brain expres
c 712	16	1.9	72928	21	AAA80253	BAC containing rep	c 785	15	1.8	201	22	AAK46186	Human bone marrow
c 713	16	1.9	85680	21	AAF22299	BAC containing rep	c 786	15	1.8	201	22	AAI52093	Probe #15490 for g
c 714	16	1.9	85684	21	AAF22292	BAC containing rep	c 787	15	1.8	201	22	AAI52093	Probe #20779 used
c 715	16	1.9	91507	23	ABL12300	Drosophila melanog	c 788	15	1.8	201	24	ABS20496	Human genome-deriv
c 716	16	1.9	95223	21	AAF22282	BAC containing rep	c 789	15	1.8	203	23	AAH84786	DNA encoding novel
c 717	16	1.9	111309	20	AAH20250	Borrelia burgdorfe	c 790	15	1.8	212	25	AAQ72529	Osteoclast-specifi
c 718	16	1.9	116624	19	AAV52850	Human eyal gene co	c 791	15	1.8	212	24	ABN86718	Human osteoclast-s
c 719	16	1.9	126512	24	ABN83429	BAC transporter	c 792	15	1.8	219	18	AAH67425	H. pylori membrane
c 720	16	1.9	134499	21	AAF22286	BAC containing rep	c 793	15	1.8	223	22	AAH25070	Human ovarian PCR-
c 721	16	1.9	147419	24	ABK83574	Human cDNA differe	c 794	15	1.8	228	21	AAH15962	Human secreted pro
c 722	16	1.9	151826	21	AAF22291	BAC containing rep	c 795	15	1.8	231	24	ABK80678	Bacillus clausii g
c 723	16	1.9	160771	24	ABQ88179	Human osteoblast d	c 796	15	1.8	231	24	ABK80944	Bacillus clausii g
c 724	16	1.9	163319	21	AAF22306	Arabidopsis thalia	c 797	15	1.8	233	16	AAH23966	Human gene signatu
c 725	16	1.9	168575	22	AAH21613	Human hypcretin r	c 798	15	1.8	235	24	ABQ73033	Human chloride cha
c 726	16	1.9	265118	22	AAH41227	Pyrococcus abyssi	c 799	15	1.8	251	21	AAA40245	Pig heart ischaemi
c 727	16	1.9	465237	24	ABQ87681	Human oestrogen re	c 800	15	1.8	255	22	AAH81997	Rat differential t
c 728	16	1.9	465237	24	ABA90193	Human oestrogen re	c 801	15	1.8	256	21	AAH91723	Mouse SmlIM/CRP2 g
c 729	16	1.9	513445	22	AAI61373	Soybean 380L3 reg	c 802	15	1.8	256	21	AAH64717	Introm 2 fragment
c 730	16	1.9	611590	21	AAF22303	Arabidopsis thalia	c 803	15	1.8	263	24	ABL67103	Thyroid cancer rel
c 731	16	1.9	910715	20	AAH20248	Borrelia burgdorfe	c 804	15	1.8	265	24	AAH21352	Human gene signatu
c 732	16	1.9	1082138	21	AAF22305	Arabidopsis thalia	c 805	15	1.8	265	16	AAH21352	Human secreted pro
c 733	16	1.9	1163020	24	ABQ67197	Listeria innocua c	c 806	15	1.8	268	21	AAH01187	Human chloride cha
c 734	16	1.9	1230025	20	AAH91990	Nucleotide sequenc	c 807	15	1.8	270	24	ABQ73040	Human chloride cha
c 735	16	1.9	1503900	22	AAH95240	Human neuregulin-1	c 808	15	1.8	273	23	AAH49216	Staphylococcus aur
c 736	16	1.9	1503900	22	AAH95240	Human neuregulin-1	c 809	15	1.8	276	21	AAH18433	Human secreted pro
c 737	16	1.9	1503900	22	AAH96733	Human neuregulin-1	c 810	15	1.8	277	21	AAA60447	Murine factor V 3'
c 738	16	1.9	1503900	22	AAH96733	Human neuregulin-1	c 811	15	1.8	277	21	AAH27771	Human secreted pro
c 739	16	1.9	1664976	19	AAV21209	Methanococcus jann	c 812	15	1.8	277	21	AAH31276	Human secreted pro



959 15 1.8 474 22 AAI92330 Human polynucleoti  
 c 960 15 1.8 474 22 AAK01925 Human brain expres  
 c 961 15 1.8 474 22 AAK27382 Human bone marrow  
 c 962 15 1.8 474 22 AAI11958 Probe #1891 for ge  
 c 963 15 1.8 474 22 AAI33289 Probe #1895 used t  
 c 964 15 1.8 474 22 AAI01890 Probe #1881 used t  
 c 965 15 1.8 474 22 AAI01885 Human genome-deriv  
 c 966 15 1.8 474 22 ABO04729 Human genome-deriv  
 c 967 15 1.8 475 24 ABL80100 Human ovarian can  
 c 968 15 1.8 476 23 ABA45314 Human prostate exp  
 c 969 15 1.8 479 22 ABA45380 Human breast cell  
 c 970 15 1.8 479 22 ABA55870 Human foetal liver  
 c 971 15 1.8 479 22 ABA25544 Probe #4010 for ge  
 c 972 15 1.8 479 22 AAI99616 Human expressed po  
 c 973 15 1.8 479 22 AAK04088 Human brain expres  
 c 974 15 1.8 479 22 AAI14144 Probe #4077 for ge  
 c 975 15 1.8 479 22 AAI35529 Probe #4215 used t  
 c 976 15 1.8 479 22 AAI03992 Probe #3983 used t  
 c 977 15 1.8 479 22 AAI03992 DNA encoding human  
 c 978 15 1.8 479 22 AAS33124 Human prostate exp  
 c 979 15 1.8 479 23 ABV48494 Human genome-deriv  
 c 980 15 1.8 482 23 ABV52493 Human prostate exp  
 c 981 15 1.8 482 24 ABK45812 cDNA encoding colo  
 c 982 15 1.8 483 23 ABV59218 Human prostate exp  
 c 983 15 1.8 485 22 ABA46438 Human breast cell  
 c 984 15 1.8 485 22 ABA57033 Human foetal liver  
 c 985 15 1.8 485 24 AAD39217 Rice receptor prot  
 c 986 15 1.8 486 22 ABA59242 Human foetal liver  
 c 987 15 1.8 486 22 ABA27987 Probe #6453 for ge  
 c 988 15 1.8 486 22 AAK07451 Human brain expres  
 c 989 15 1.8 486 22 AAK33233 Human bone marrow  
 c 990 15 1.8 486 22 AAI16376 Probe #6309 for ge  
 c 991 15 1.8 486 22 AAI39028 Probe #7714 used t  
 c 992 15 1.8 486 22 AAH90785 CPE 89 coding sequ  
 c 993 15 1.8 486 23 AAS55628 Streptococcus pneu  
 c 994 15 1.8 486 23 AAS55880 Streptococcus pneu  
 c 995 15 1.8 486 24 ABO08064 Human genome-deriv  
 c 996 15 1.8 487 21 AAC46720 Zea mays DNA fragm  
 c 997 15 1.8 490 22 AAK66769 Human immune/haema  
 c 998 15 1.8 490 22 AAK66770 Human immune/haema  
 c 999 15 1.8 494 23 ABV49671 Human prostate exp  
 1000 15 1.8 499 22 AAI92101 Human polynucleoti

## ALIGNMENTS

RESULT 1  
 ABK68876  
 ID ABK68876 standard; DNA; 840 BP.  
 XX AC ABK68876;  
 XX  
 DT 02-JUL-2002 (first entry)  
 XX  
 DE DNA encoding Ehrlichia canis p28-2.  
 XX  
 KW Ehrlichia canis infection; vaccine; serodiagnostic; gene; p28; ss;  
 KW antibacterial.  
 XX  
 OS Ehrlichia canis.  
 XX  
 PN WO20022782-A2.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PF 12-SEP-2001; 2001WO-US28759.  
 XX  
 PR 12-SEP-2000; 2000US-0660587.  
 XX  
 PA (RERE-) RES DEV FOUND.  
 XX  
 PI Walker DH, Yu X, McBride JW;

XX WPI; 2002-351882/38.  
 DR P-PSDB; AAU96116.  
 XX  
 PT New recombinant homologous 28 kilodalton immunodominant protein from  
 PT Ehrlichia canis, useful for treating Ehrlichia canis infections -  
 XX Claim 5; Figure 14; 106pp; English.  
 XX  
 CC The invention relates to a recombinant homologous 28 kDa immunodominant  
 CC protein, p28 (I) of Ehrlichia canis. (I), a 28-kDa antigen preferably  
 CC dispersed in a pharmaceutically acceptable carrier, is useful for  
 CC inhibiting E. canis infection in a subject. (I) is useful in the  
 CC development of vaccines and serodiagnostics that are particularly  
 CC effective for disease prevention and serodiagnosis. ABK68852-ABK68878  
 CC represent the 28-kDa antigen coding sequences and PCR primers of the  
 CC invention.  
 XX Sequence 840 BP; 301 A; 141 C; 144 G; 254 T; 0 other;  
 SQ  
 Query Match 100.0%; Score 840; DB 24; Length 840;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGAATTATAAGAAAATTTCTAGTAAGAGCGGTTAATCTCATTAAATCTCAATCTTACCA 60  
 DB 1 ATGAATTATAAGAAAATTTCTAGTAAGAGCGGTTAATCTCATTAAATCTCAATCTTACCA 60  
 QY 61 TATCAGTCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGATAACAAGAGGCTTTCTAC 120  
 DB 61 TATCAGTCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGATAACAAGAGGCTTTCTAC 120  
 QY 121 ATTAGTCAAAAGTACAATCCAAGTATATACACATTTAGAAAATTTCTCGCTGAAGAACT 180  
 DB 121 ATTAGTCAAAAGTACAATCCAAGTATATACACATTTAGAAAATTTCTCGCTGAAGAACT 180  
 QY 181 CCTATTATGAACAAATTTCTCACTAAAAAGTTTTCGGACTAAAGAGAGTGGTGTAT 240  
 DB 181 CCTATTATGAACAAATTTCTCACTAAAAAGTTTTCGGACTAAAGAGAGTGGTGTAT 240  
 QY 241 ATAACAAAAAAGACGATTTTACAAGAGTAGCTCCAGGCATTGATTTTCAAAAATAACTTA 300  
 DB 241 ATAACAAAAAAGACGATTTTACAAGAGTAGCTCCAGGCATTGATTTTCAAAAATAACTTA 300  
 QY 301 ATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGGACCGACCAAGATAGAACTTCAA 360  
 DB 301 ATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGGACCGACCAAGATAGAACTTCAA 360  
 QY 361 GCTGCATATCAACAAATTTAATCCAAAAACACCGATAACAATGATCTACTGATAATGGTCAA 420  
 DB 361 GCTGCATATCAACAAATTTAATCCAAAAACACCGATAACAATGATCTACTGATAATGGTCAA 420  
 QY 421 TACTATAAACAATTTTGCATTAATCTCGTAAAGATGCAATGGAAGATCAGCAATATAGTA 480  
 DB 421 TACTATAAACAATTTTGCATTAATCTCGTAAAGATGCAATGGAAGATCAGCAATATAGTA 480  
 QY 481 CTTAAAAATGACGCATAAATTTTATGTCATTGATGGTTAATCTCTGCTATGACATTTACA 540  
 DB 481 CTTAAAAATGACGCATAAATTTTATGTCATTGATGGTTAATCTCTGCTATGACATTTACA 540  
 QY 541 GCTGAAGGAGTATCTTTCTGACCATATGTCATGTCAGGTATAGGAGCAGATCTTATCACT 600  
 DB 541 GCTGAAGGAGTATCTTTCTGACCATATGTCATGTCAGGTATAGGAGCAGATCTTATCACT 600  
 QY 601 ATTTTAAAGACCTCAATCTAAAAATTTCTTACCAAGAAAAAATAGGTATTAGTTACCCCT 660  
 DB 601 ATTTTAAAGACCTCAATCTAAAAATTTCTTACCAAGAAAAAATAGGTATTAGTTACCCCT 660  
 QY 661 ATCACACCAGAGTCTCTGCAATTTATTTGGTGATACCTACCATGGCGTTATTGGTAAATAA 720  
 DB 661 ATCACACCAGAGTCTCTGCAATTTATTTGGTGATACCTACCATGGCGTTATTGGTAAATAA 720  
 QY 721 TTTGAGAAGATACCTGTAAATAACTCTCTAGTATTAAATGATGCTCTCTCAAACCATCTCT 780

```
Db 721 TTTGAGAGATACCTGTAAATACCTCTGTAGTATTAAATGATGCTCTCCAAACCACTCT 780
Oy 781 GCTTCAGTAACCTTTGAGCTTGGTATCTTTGGCGAGAAATTTGGAATGAGGTTTCACTTC 840
Db 781 GCTTCAGTAACCTTTGAGCTTGGTATCTTTGGCGAGAAATTTGGAATGAGGTTTCACTTC 840

RESULT 2
AAX34770
ID AAX34770 standard; DNA; 495 BP.
AC AAX34770;
XX
DT 05-JUL-1999 (first entry)
DE DNA encoding P30-10protein.
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog; ss.
XX Ehrlichia canis.
XX OS
XX WO9913720-A1.
XX
XX 25-MAR-1999.
XX
XX 18-SEP-1998; 98WO-US19600.
XX
XX 19-SEP-1997; 97US-0059353.
XX
XX (OHIS ) UNIV OHIO STATE.
XX
XX Ohashi N, Rikihisa Y;
XX WPI; 1999-254290/21.
XX P-PSDB; AAY06970.
XX
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
XX Ehrlichia canis
XX
XX Disclosure; Fig 30A; 55pp; English.
XX
XX The invention provides isolated outer membrane proteins (OMP) from
XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
XX of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
XX in AAY06943-958. The E. canis proteins form part of the P30 family and
XX consist of proteins shown in AAY06959-970. The proteins and genes are
XX used to detect E. chaffeensis in patients and E. canis in dogs.
XX
XX Sequence 495 BP; 192 A; 81 C; 80 G; 142 T; 0 other;
XX
Query Match 44.2%; Score 371; DB 20; Length 495;
Best Local Similarity 100.0%; Pred. No. 6e-174;
Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 ATGAATTATAGAAAATTTCTAGTAGAGCGGTTAATCTCATTAATGCTCAATCTTACCA 60
Db 1 ATGAATTATAGAAAATTTCTAGTAGAGCGGTTAATCTCATTAATGCTCAATCTTACCA 60
Oy 61 TATCAGCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120
Db 61 TATCAGCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120
Oy 121 ATTAGTCAAAAGTACAATCCCAAGTATATCACATTTAGAAAATTTCTGCTGAGAACT 180
Db 121 ATTAGTCAAAAGTACAATCCCAAGTATATCACATTTAGAAAATTTCTGCTGAGAACT 180
Oy 181 CCTATTATGGAACAAATTTCTCTCACTAAAAAGTTTTTCGGACTAAAGAAAGATGGTGAT 240
Db 181 CCTATTATGGAACAAATTTCTCTCACTAAAAAGTTTTTCGGACTAAAGAAAGATGGTGAT 240
Oy 241 ATAACAAAAAAGAGGATTTTACAGAGTAGTCTCCAGGCATTTGATTTTCAAAATAACTTA 300
```

```
Db 241 ATAACAAAAAAGAGGATTTTACAGAGTAGTCTCCAGGCATTTGATTTTCAAAATAACTTA 300
Oy 301 ATATCAGGATTTTTCAGGAAGTATTGGTTACTCTATGACGCGACCAAGAAATAGAACTTGAA 360
Db 301 ATATCAGGATTTTTCAGGAAGTATTGGTTACTCTATGACGCGACCAAGAAATAGAACTTGAA 360
Oy 361 GCTGCATATCA 371
Db 361 GCTGCATATCA 371

RESULT 3
AAX34744
ID AAX34744 standard; DNA; 852 BP.
XX
XX AC AAX34744;
XX
XX 05-JUL-1999 (first entry)
DE DNA encoding OMP-1B protein.
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog; ss.
XX
XX Ehrlichia chaffeensis.
XX OS
XX WO9913720-A1.
XX
XX 25-MAR-1999.
XX
XX 18-SEP-1998; 98WO-US19600.
XX
XX 19-SEP-1997; 97US-0059353.
XX
XX (OHIS ) UNIV OHIO STATE.
XX
XX Ohashi N, Rikihisa Y;
XX WPI; 1999-254290/21.
XX P-PSDB; AAY06944.
XX
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
XX Ehrlichia canis
XX
XX Disclosure; Fig 4A; 55pp; English.
XX
XX The invention provides isolated outer membrane proteins (OMP) from
XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
XX of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
XX in AAY06943-958. The E. canis proteins form part of the P30 family and
XX consist of proteins shown in AAY06959-970. The proteins and genes are
XX used to detect E. chaffeensis in patients and E. canis in dogs.
XX
XX Sequence 852 BP; 301 A; 155 C; 157 G; 239 T; 0 other;
XX
Query Match 3.6%; Score 30; DB 20; Length 852;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 523 ACTTGCTATGACATTACAGCTGAAGGAGTA 552
Db 532 ACTTGCTATGACATTACAGCTGAAGGAGTA 561

RESULT 4
ABK68873/c
ID ABK68873 standard; DNA; 22 BP.
XX
XX AC ABK68873;
XX
XX 02-JUL-2002 (first entry)
XX
```

DE Ehrlichia canis p28-5 gene, PCR primer Eca5'-1600.  
XX  
XX Ehrlichia canis infection; vaccine; serodiagnostic; PCR; primer; p28; ss;  
KW antibacterial.  
XX  
XX Ehrlichia canis.  
OS Synthetic.  
OS  
XX WO200222782-A2.  
PN  
XX 21-MAR-2002.  
PD  
XX 12-SEP-2001; 2001WO-US28759.  
PF  
XX 12-SEP-2000; 2000US-0660587.  
PR  
XX (RERE-) RES DEV FOUND.  
PA  
XX Walker DH, Yu X, McBride JW;  
PI WPI; 2002-351882/38.  
XX  
XX New recombinant homologous 28 kilodalton immunodominant protein from  
PT Ehrlichia canis, useful for treating Ehrlichia canis infections  
PT  
XX Example 8; Page 42; 106pp; English.  
PS  
XX The invention relates to a recombinant homologous 28 kDa immunodominant  
CC protein, p28 (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably  
CC dispersed in a pharmaceutically acceptable carrier, is useful for  
CC inhibiting E. canis infection in a subject. (I) is useful in the  
CC development of vaccines and serodiagnostics that are particularly  
CC effective for disease prevention and serodiagnosis. ABK68852-ABK68878  
CC represent the 28-kDa antigen coding sequences and PCR primers of the  
CC invention.  
CC  
XX Sequence 22 BP; 9 A; 6 C; 3 G; 4 T; 0 other;  
SQ

Query Match 2.6%; Score 22; DB 24; Length 22;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 670 GAAGTCTCTGCATTATTGGTG 691  
DB 22 GAAGTCTCTGCATTATTGGTG 1

RESULT 5  
AAC14373  
ID AAC14373 standard; cDNA; 206 BP.  
XX  
XX AAC14373;  
AC  
XX 06-OCT-2000 (first entry)  
DT  
XX Human secreted protein 5' EST, SEQ ID NO: 18448.  
DE  
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.  
KW  
XX Homo sapiens.  
OS  
XX EP1033401-A2.  
PN  
XX 06-SEP-2000.  
PD  
XX 21-FEB-2000; 2000EP-0200610.  
PF  
XX 26-FEB-1999; 99US-0122487.  
PR  
XX (GEST ) GENSET.  
PA  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
PI

XX WPI; 2000-500381/45.  
DR  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
PT  
XX Claim 1; SEQ ID 18448; 71pp + CD-ROM; English.  
PS  
XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dr primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.  
XX  
XX Sequence 206 BP; 50 A; 34 C; 37 G; 85 T; 0 other;  
SQ

Query Match 2.3%; Score 19; DB 21; Length 206;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 TTTTGCAGATCCTGTAGGT 87  
DB 34 TTTTGCAGATCCTGTAGGT 52

RESULT 6  
AAD01294  
ID AAD01294 standard; DNA; 840 BP.  
XX  
XX AAD01294;  
AC  
XX 12-OCT-2000 (first entry)  
DT  
XX Ehrlichia canis immunoreactive protein Eca28SA3 DNA.  
DE  
XX Homologous 28-kDa protein gene; Eca28SA3; immunoreactive; vaccine;  
KW p28 gene; polymorphic multiple gene family; immunoprotective antigen;  
KW antibacterial; canine ehrlichiosis; canine tropical pancytopenia;  
KW tick-borne rickettsial disease; serodiagnosis; ds.  
XX  
XX Ehrlichia canis.  
OS

Key Location/Qualifiers  
FH 1..840  
CDS /tag= a  
FT /product= "Eca28SA3 protein (30-kDa)"  
FT /note= "Does not include stop codon"  
FT /partial  
FT sig\_peptide 1..69  
FT /tag= b  
FT mat\_peptide 70..840  
FT /tag= c  
FT /product= "Mature Eca28SA3 protein (28-kDa)"  
XX  
XX WO200032745-A2.  
PN  
XX 08-JUN-2000.  
PD  
XX 24-NOV-1999; 99WO-US28075.  
PF  
XX 30-NOV-1998; 98US-0201458.  
PR  
XX 03-MAR-1999; 99US-0261358.  
PR  
XX

PA (RERE-) RES- DEV FOUND.  
 XX Walker DH, Yu X, McBride JW;  
 PI  
 XX WPI; 2000-412298/35.  
 DR P-PSDB; AAU71479.  
 XX Ehrlichia canis antigens useful for vaccinating against canine  
 PT ehrlichiosis in dogs -  
 XX  
 XX Claim 5; Page 67-68; 86pp; English.  
 PS The patent relates to homologous 28-kilodalton (kDa) protein genes of  
 CC Ehrlichia canis, designated Eca28SA1, Eca28SA2, Eca28SA3, Eca28-1 and  
 CC Eca28-2. These genes are members of a polymorphic multiple gene family  
 CC and contained in a single locus of 5.592 kb. The 28-kDa proteins are  
 CC immunoreactive with anti-E. canis serum hence are important  
 CC immunoprotective antigens. The protein is useful for vaccinating  
 CC against E. canis infections such as canine ehrlichiosis in dogs.  
 CC Canine ehrlichiosis, also known as canine tropical pancytopenia, is a  
 CC tick-borne rickettsial disease of dogs. Eca28-1 is conserved amongst  
 CC different strains of E. canis and hence useful for serodiagnosis of  
 CC canine ehrlichiosis. The present sequence is a DNA encoding E. canis  
 CC Eca28SA3 30-kDa protein which is post-translationally modified to a  
 CC mature 28-kDa protein by cleavage of N-terminal signal sequence.  
 XX  
 SQ Sequence 840 BP; 282 A; 137 C; 150 G; 271 T; 0 other;  
 Query Match 2.3%; Score 19; DB 21; Length 840;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 343 CCAAGAATAGAACTTGAAG 361  
 DB 337 CCAAGAATAGAACTTGAAG 355  
 RESULT 7  
 ABK68854  
 ID ABK68854 standard; DNA; 840 BP.  
 XX  
 AC ABK68854;  
 XX  
 DT 02-JUL-2002 (first entry)  
 DE DNA encoding Ehrlichia canis p28-6.  
 XX  
 KW Ehrlichia canis infection; vaccine; serodiagnostic; gene; p28; ss;  
 KW antibacterial.  
 XX  
 OS Ehrlichia canis.  
 XX  
 PN WO200222782-A2.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PF 12-SEP-2001; 2001WO-US28759.  
 XX  
 PR 12-SEP-2000; 2000US-0660587.  
 XX  
 PA (RERE-) RES DEV FOUND.  
 XX Walker DH, Yu X, McBride JW;  
 PI  
 XX WPI; 2002-351882/38.  
 DR P-PSDB; AAU96102.  
 XX  
 PT New recombinant homologous 28 kilodalton immunodominant protein from  
 PT Ehrlichia canis, useful for treating Ehrlichia canis infections -  
 XX  
 PS Claim 5; Figure 7; 106pp; English.  
 XX  
 CC The invention relates to a recombinant homologous 28 kDa immunodominant

CC protein, p28 (I) of Ehrlichia canis. (I), a 28-kDa antigen preferably  
 CC dispersed in a pharmaceutically acceptable carrier, is useful for  
 CC inhibiting E. canis infection in a subject. (I) is useful in the  
 CC development of vaccines and serodiagnostics that are particularly  
 CC effective for disease prevention and serodiagnosis. ABK68852-ABK68878  
 CC represent the 28-kDa antigen coding sequences and PCR primers of the  
 CC invention.  
 XX  
 SQ Sequence 840 BP; 282 A; 137 C; 150 G; 271 T; 0 other;  
 Query Match 2.3%; Score 19; DB 24; Length 840;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 343 CCAAGAATAGAACTTGAAG 361  
 DB 337 CCAAGAATAGAACTTGAAG 355  
 RESULT 8  
 AAX34762  
 ID AAX34762 standard; DNA; 843 BP.  
 XX  
 AC AAX34762;  
 XX  
 DT 05-JUL-1999 (first entry)  
 DE DNA encoding P30-2 protein.  
 XX  
 KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 KW detection; dog; ss.  
 XX  
 OS Ehrlichia canis.  
 XX  
 PN WO9913720-A1.  
 XX  
 PD 25-MAR-1999.  
 XX  
 PF 18-SEP-1998; 98WO-US19600.  
 XX  
 PR 19-SEP-1997; 97US-0059353.  
 XX  
 PA (OHIS ) UNIV OHIO STATE.  
 XX  
 PI Ohashi N, Rikihisa Y;  
 XX  
 DR WPI; 1999-254290/21.  
 DR P-PSDB; AAY06962.  
 XX  
 PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
 PT Ehrlichia canis  
 XX  
 PS Disclosure; Fig 22A; 55pp; English.  
 XX  
 CC The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and  
 CC consist of proteins shown in AAY06959-970. The proteins and genes are  
 CC used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX  
 SQ Sequence 843 BP; 283 A; 136 C; 150 G; 274 T; 0 other;  
 Query Match 2.3%; Score 19; DB 20; Length 843;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 343 CCAAGAATAGAACTTGAAG 361  
 DB 337 CCAAGAATAGAACTTGAAG 355  
 RESULT 9

```

AA34749
ID AAX34749 standard; DNA; 894 BP.
XX AC AAX34749;
XX DT 05-JUL-1999 (first entry)
XX DE DNA encoding OMP-1A protein.
XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
XX KW detection; dog; ss.
XX OS Ehrlichia chaffeensis.
XX PN WO9913720-A1.
XX PD 25-MAR-1999.
XX PF 18-SEP-1998; 98WO-US19600.
XX PR 19-SEP-1997; 97US-0059353.
XX PA (OHIS ) UNIV OHIO STATE.
XX PI Ohashi N, Rikihisa Y;
XX WPI; 1999-254290/21.
XX DR P-PSDB; AAY06949.
XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and
XX PT Ehrlichia canis
XX PS Disclosure; Fig 9A; 55pp; English.
XX CC The invention provides isolated outer membrane proteins (OMP) from
XX CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
XX CC of the OMP family and consist of proteins OMP-1, -(B to Z) shown
XX CC in AAY06943-958. The E. canis proteins form part of the P30 family and
XX CC consist of proteins shown in AAY06959-970. The proteins and genes are
XX CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX SQ Sequence 894 BP; 308 A; 143 C; 152 G; 291 T; 0 other;

Query Match 2.3%; Score 19; DB 20; Length 894;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 620 TAAATTTGCTTACCAAGG 638
DB 671 TAAATTTGCTTACCAAGG 689

RESULT 10
AAD01295
ID AAD01295 standard; DNA; 2037 BP.
XX AC AAD01295;
XX DT 12-OCT-2000 (first entry)
XX DE Ehrlichia canis immunoreactive protein genes ECA28SA2 and ECA28SA3.
XX KW Homologous mature 28-kDa protein gene; ECA28SA2; ECA28SA3; vaccine;
XX KW p28 gene; polymorphic multiple gene family; immunoprotective antigen;
XX KW antibacterial; canine ehrlichiosis; canine tropical pancytopenia;
XX KW tick-borne rickettsial disease; serodiagnosis; immunoreactive; ds.
XX OS Ehrlichia canis.
XX FH Key Location/Qualifiers
XX FT CDS 1..852
XX FT /tag= a
XX FT /product= "ECA28SA2 protein (30-kDa)"

```

```

FT misc_feature 853..1194
FT /tag= b
FT /note= "Intergenic non-coding region NC2"
FT CDS 1195..2037
FT /tag= c
FT /product= "ECA28SA3 protein (30-kDa)"
FT sig_peptide 1195..1263
FT /tag= d
FT mat_peptide 1264..2034
FT /tag= e
FT /product= "Mature ECA28SA3 protein (28-kDa)"
XX PN WO200032745-A2.
XX PD 08-JUN-2000.
XX PF 24-NOV-1999; 99WO-US28075.
XX PR 30-NOV-1998; 98US-0201458.
XX PR 03-MAR-1999; 99US-0261358.
XX PA (RERE-) RES DEV FOUND.
XX PI Walker DH, Yu X, McBride JW;
XX WPI; 2000-412298/35.
XX DR P-PSDB; AAY71478, AAY71479.
XX PT Ehrlichia canis antigens useful for vaccinating against canine
XX PT ehrlichiosis in dogs -
XX PS Example 15; Fig 7; 86pp; English.
XX CC The patent relates to homologous 28-kiloDalton (kDa) protein genes of
XX CC Ehrlichia canis, designated ECA28SA1, ECA28SA2, ECA28SA3, ECA28-1 and
XX CC ECA28-2. These genes are members of a polymorphic multiple gene family
XX CC and contained in a single locus of 5,592 kb. The 28-kDa proteins are
XX CC immunoreactive with anti-E. canis serum hence are important
XX CC immunoprotective antigens. The protein is useful for vaccinating
XX CC against E. canis infections such as canine ehrlichiosis in dogs.
XX CC Canine ehrlichiosis, also known as canine tropical pancytopenia, is a
XX CC tick-borne rickettsial disease of dogs. ECA28-1 is conserved amongst
XX CC different strains of E. canis and hence useful for serodiagnosis of
XX CC canine ehrlichiosis. The present sequence is that of homologous
XX CC genes encoding E. canis ECA28SA2 and ECA28SA3 30-kDa proteins which
XX CC are post-translationally modified to corresponding
XX CC mature 28-kDa proteins by cleavage of N-terminal signal sequence.
XX SQ Sequence 2037 BP; 687 A; 321 C; 342 G; 687 T; 0 other;

Query Match 2.3%; Score 19; DB 21; Length 2037;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 343 CCAAGATAGAACTTGAAG 361
DB 1531 CCAAGATAGAACTTGAAG 1549

RESULT 11
AAS45313/c
ID AAS45313 standard; DNA; 5979 BP.
XX AC AAS45313;
XX DT 18-DEC-2001 (first entry)
XX DE Chemically pretreated complementary DNA associated with cell cycle #9.
XX KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
XX KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
XX KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
XX KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;

```



```

KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW PCR primer.
XX
XX Homo sapiens.
OS
XX WO200168911-A2.
PN
XX
XX 20-SEP-2001.
PD
XX
XX 15-MAR-2001; 2001WO-EP02945.
PF
XX 15-MAR-2000; 2000DE-1013847.
PR
XX 06-APR-2000; 2000DE-1019058.
PR
XX 07-APR-2000; 2000DE-1019173.
PR
XX 30-JUN-2000; 2000DE-1032529.
PR
XX 01-SEP-2000; 2000DE-1043826.
PR
XX
XX (EPIG-) EPIGENOMICS AG.
PA
XX
XX Olek A, Piepenbrock C, Berlin K;
PI
XX
XX WPI; 2001-602751/68.
DR
XX
XX Designing primers and probes for analysing diseases associated with
PT cytosine methylation state e.g. arthritis, cancer, aging.
PT arteriosclerosis comprising fragments of chemically modified genes
PT associated with cell cycle.
XX
XX Claim 1; SEQ ID No 18; 28pp; English.
PS
XX
XX Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC molecules associated with the cell cycle and specific PCR primers of the
CC invention. The sequences are useful for detecting the methylation state
CC of all CpG dinucleotides in a sequence and therefore for analysing
CC associated diseases. By analysing cytosine methylations in the pretreated
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC of existing diseases or the predisposition to specific diseases can be
CC ascertained. The parameters may be compared to another set of genetic
CC and/or epigenetic parameters, the differences serving as basis for
CC diagnosis and/or prognosis events which are disadvantageous to patients.
CC The sequences of the invention are useful for the diagnosis and therapy
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC aging, glomerular disease, Lewy body disease, arthritis,
CC arteriosclerosis, solid tumours and cancers.
XX
XX Sequence 5979 BP; 2294 A; 42 C; 812 G; 2831 T; 0 other;
SQ
Query Match 2.3%; Score 19; DB 22; Length 5979;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 285 TTTTCAAAATAACTTAATA 303
Db 1203 TTTTCAAAATAACTTAATA 1185
RESULT 12
ABK28152/c
ID ABK28152 standard; DNA; 5979 BP.
XX
XX AC ABK28152;
XX
XX 23-APR-2002 (first entry)
XX
XX DNA transcription associated complementary genomic DNA #13.
XX
XX DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW immunological disorder; Werner syndrome; developmental disorder;
KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW
KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW polyglutamine disorder; solid tumour.
XX
XX Unidentified.
OS
XX WO200192565-A2.
PN
XX
XX 06-DEC-2001.
PD
XX
XX 06-APR-2001; 2001WO-EP03973.
PF
XX
XX 06-APR-2000; 2000DE-1019058.
PR
XX 07-APR-2000; 2000DE-1019173.
PR
XX 30-JUN-2000; 2000DE-1032529.
PR
XX 01-SEP-2000; 2000DE-1043826.
PR
XX
XX (EPIG-) EPIGENOMICS AG.
PA
XX
XX Olek A, Piepenbrock C, Berlin K;
PI
XX
XX WPI; 2002-090046/12.
DR
XX
XX New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological
PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid
PT tumours or cancer.
XX
XX Claim 1; SEQ ID No 26; 32pp; English.
PS
XX
XX The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC neurocrosis, developmental disorders, psoriasis, Rieger's syndrome,
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX Sequence 5979 BP; 2294 A; 42 C; 812 G; 2831 T; 0 other;
SQ
Query Match 2.3%; Score 19; DB 24; Length 5979;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 285 TTTTCAAAATAACTTAATA 303
Db 1203 TTTTCAAAATAACTTAATA 1185
RESULT 13
AAS46343/c
ID AAS46343 standard; DNA; 6179 BP.
XX
XX AC AAS46343;
XX
XX 18-DEC-2001 (first entry)
XX
XX Tumour suppressor gene derived chemically modified sequence #65.
XX

```

```
XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX Homo sapiens.
XX WO200168912-A2.
XX 20-SEP-2001.
XX 15-MAR-2001; 2001WO-EP02955.
XX 15-MAR-2000; 2000DE-1013847.
XX 06-APR-2000; 2000DE-1019058.
XX 07-APR-2000; 2000DE-1019173.
XX 30-JUN-2000; 2000DE-1032529.
XX 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2001-602752/68.
XX Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT analysing diseases associated with cytosine methylation state e.g.
PT cancer.
XX Claim 1; SEQ ID No 65; 27pp; English.
XX The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and
CC oncogenes having a sequence taken from 536 (actually 533 since
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC (SS) and sequences complementary to (SS). The nucleic acid may be a
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC form part of a set of probes for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms and also to be used in an
CC array for analysing diseases associated with CpG dinucleotides e.g.
CC cancers and tumours. The probes can also be used in a method for
CC ascertaining genetic and/or epigenetic parameters for the diagnosis
CC and/or therapy of existing diseases or the predisposition to specific
CC diseases, by analysing cytosine methylations. The parameters may be
CC compared to another set of genetic and/or epigenetic parameters, the
CC differences serving as basis for diagnosis and/or prognosis events which
CC are disadvantageous to patients. The present sequence is one of the
CC 533 genomic sequences derived from tumour suppressor genes and
CC oncogenes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 6179 BP; 1761 A; 105 C; 1302 G; 3011 T; 0 other;
SQ Query Match 2.3%; Score 19; DB 22; Length 6179;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 284 ATTTCAAAATAACTTAAT 302
DB 2746 ATTTCAAAATAACTTAAT 2728
RESULT 14
ABK31250/c
ID ABK31250 standard; DNA; 6179 BP.
XX AC ABK31250;
XX
```

```
DT 23-APR-2002 (first entry)
XX Signal transduction associated gene modified DNA #47.
XX Human; signal transduction associated gene; cytosine methylation state;
KW CpG island; signal transduction associated disease; solid tumour; cancer;
KW antitumour; cytostatic; mutant; ds.
XX Homo sapiens.
XX OS Synthetic.
XX WO200200926-A2.
XX 03-JAN-2002.
XX 29-JUN-2001; 2001WO-EP07472.
XX 30-JUN-2000; 2000DE-1032529.
XX 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-147896/19.
XX Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with signal transduction.
XX Claim 1; SEQ ID No 93; 24pp; English.
XX The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically
CC modified using a solution of bisulphite, hydrogen sulphite or
CC disulphite. Also disclosed are oligonucleotides and/or PNA oligomers
CC for detecting the cytosine methylation state (CpG islands) of these
CC genes, and a method for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with signal transduction.
CC The genomic DNA can be obtained from cells or cellular components which
CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,
CC cerebral-spinal fluid, tissue embedded in paraffin such as tissue from
CC eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,
CC histologic object slides, and all their possible combinations. The
CC sequences of the invention are useful for the diagnosis and therapy of
CC diseases associated with signal transduction e.g. solid tumours and
CC cancer. ABK31158-ABK31545 represent chemically pretreated genomic DNA
CC sequences of different genes associated with signal transduction, or
CC their complementary sequences.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX Sequence 6179 BP; 1761 A; 105 C; 1302 G; 3011 T; 0 other;
SQ Query Match 2.3%; Score 19; DB 24; Length 6179;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 284 ATTTCAAAATAACTTAAT 302
DB 2746 ATTTCAAAATAACTTAAT 2728
RESULT 15
ABN64597
ID ABN64597 standard; cDNA; 300 BP.
XX AC ABN64597;
XX 28-JUN-2002 (first entry)
XX Human cancer related polynucleotide SEQ ID NO 4564.
DE
```

```
XX Human; cytostatic; gene expression; gene mapping; tissue profiling;
KW gene therapy; cancer; tumour; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200214500-A2.
XX
XX 21-FEB-2002.
PD
XX
XX 16-AUG-2001; 2001WO-US25840.
PF
XX
XX 16-AUG-2000; 2000US-226326P.
PR
XX (CHIR ) CHIRON CORP.
PA
PA (HYSE-) HYSEQ INC.
XX
XX Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;
PI Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;
PI
XX WPI; 2002-241905/29.
DR
XX
XX New nucleic acid for producing a polypeptide, detecting differentially
XX expressed genes correlated with a cancerous state of a mammalian cell,
XX PT and inhibiting tumor growth -
XX
XX Claim 1; SEQ ID NO 4564; 883pp + Sequence Listing; English.
XX
XX The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
XX with cytostatic activity. The polynucleotide is used to produce a
XX polypeptide, to detect differentially expressed genes correlated with a
XX cancerous state of a mammalian cell and to inhibit tumour growth. The
XX polynucleotide is used as a probe in mapping and tissue profiling. The
XX encoded polypeptide and antibodies to the polypeptide can also be used
XX for therapeutic and diagnostic purposes. The polynucleotide is useful for
XX gene therapy.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 300 BP; 97 A; 48 C; 44 G; 111 T; 0 other;
XX
XX Query Match 2.1%; Score 18; DB 24; Length 300;
XX Best Local Similarity 100.0%; Pred. No. 94;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 425 ATAAACATTTTGCATTAT 442
XX ||||||||||||||||
XX Db 95 ATAAACATTTTGCATTAT 112
XX
XX RESULT 16
XX ABA43326/C
XX ID ABA43326 standard; DNA; 429 BP.
XX
XX AC ABA43326;
XX
XX DT 01-FEB-2002 (first entry)
XX
XX DE Human breast cell single exon nucleic acid probe #2021.
XX
XX KW Human; microarray; single exon probe; gene expression; breast;
XX disease; cancer; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200157271-A2.
XX
XX PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00662.
XX
XX PF 04-FEB-2000; 2000US-0180312.
XX
XX PR
```

```
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes,
XX useful for measuring gene expression in sample derived from human
XX breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 1; SEQ ID NO 2021; 327pp + sequence listing; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and BT 474 cells. The method involves contacting
XX the probes with a collection of detectably labelled nucleic acids
XX derived from mRNA of human breast, and then measuring the label
XX bound to each probe of the microarray. The probes are useful for
XX verifying the expression of regions of genomic DNA predicted to
XX encode proteins. They are useful for gene discovery, and for
XX determining predisposition and/or prognosing breast disease. Gene
XX expression analysis is useful for assessing the toxicity of chemical
XX agents on cells. The microarray of this invention presents a far greater
XX diversity of probes for measuring gene expression, with far less bias
XX than expressed sequence tag microarrays. The method is suitable for
XX rapid production of functional information from genomic sequence. The
XX present sequence is a single exon nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 429 BP; 148 A; 78 C; 69 G; 134 T; 0 other;
XX
XX Query Match 2.1%; Score 18; DB 22; Length 429;
XX Best Local Similarity 100.0%; Pred. No. 93;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 635 AAGGAAAAATAGGTATTA 652
XX ||||||||||||||||
XX Db 131 AAGGAAAAATAGGTATTA 114
XX
XX RESULT 17
XX ABA53769/C
XX ID ABA53769 standard; DNA; 429 BP.
XX
XX AC ABA53769;
XX
XX DT 01-FEB-2002 (first entry)
XX
XX DE Human foetal liver single exon nucleic acid probe #2074.
XX
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200157277-A2.
XX
XX PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00669.
XX
XX PF 04-FEB-2000; 2000US-0180312.
XX
XX PR 26-MAY-2000; 2000US-0207456.
XX
XX PR 30-JUN-2000; 2000US-0608408.
XX
XX PR 03-AUG-2000; 2000US-0632366.
```

```
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver.
XX
PS Claim 1; SEQ ID NO 2074; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 429 BP; 148 A; 78 C; 69 G; 134 T; 0 other;

Query Match 2.1%; Score 18; DB 22; Length 429;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 AAGGAAAAATAGGTATTA 652
DB 131 AAGGAAAAATAGGTATTA 114

RESULT 18
ABA23516/c
ID ABA23516 standard; DNA; 429 BP.
XX
AC ABA23516;
XX
XX 23-JAN-2002 (first entry)
XX
DE Probe #1982 for gene expression analysis in human heart cell sample.
XX
XX Human; gene expression; heart; microarray; vascular system; probe;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease; ss.
XX
XX Homo sapiens.
XX
OS WO200157274-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00666.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX Claim 1; SEQ ID No 1982; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX Note: the sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 429 BP; 148 A; 78 C; 69 G; 134 T; 0 other;

Query Match 2.1%; Score 18; DB 22; Length 429;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 AAGGAAAAATAGGTATTA 652
DB 131 AAGGAAAAATAGGTATTA 114

RESULT 19
AAK02030/c
ID AAK02030 standard; DNA; 429 BP.
XX
AC AAK02030;
XX
XX 05-NOV-2001 (first entry)
XX
XX Human brain expressed single exon probe SEQ ID NO: 2021.
XX
XX Human; brain expressed exon; gene expression analysis; probe;
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX Example 4; SEQ ID NO: 2021; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
```

CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.

XX SQ Sequence 429 BP; 148 A; 78 C; 69 G; 134 T; 0 other;

Query Match 2.1%; Score 18; DB 22; Length 429;

Best Local Similarity 100.0%; Pred. No. 93;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 635 AAGGAAAAATAGGTATTA 652

Db 131 AAGGAAAAATAGGTATTA 114

## RESULT 20

AAK27485/c

ID AAK27485 standard; DNA; 429 BP.

XX AC AAK27485;

XX DT 06-NOV-2001 (first entry)

XX DE Human bone marrow expressed single exon probe SEQ ID NO: 2042.

XX KW Human; bone marrow expressed exon; gene expression analysis; probe;

XX KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

XX OS Homo sapiens.

XX PN WO200157276-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00668.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488900/53.

XX PT Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human bone marrow -

XX PS Example 4; SEQ ID NO: 2042; 658pp + Sequence Listing; English.

XX CC The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX bone marrow. They can be used to measure gene expression in bone marrow  
XX samples, which may enable the improved diagnosis and treatment of cancers  
XX such as lymphoma, leukaemia and myeloma. The present sequence is one of  
XX the probes of the invention.

XX SQ Sequence 429 BP; 148 A; 78 C; 69 G; 134 T; 0 other;

## Query Match

Best Local Similarity 2.1%; Score 18; DB 22; Length 429;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 635 AAGGAAAAATAGGTATTA 652

Db 131 AAGGAAAAATAGGTATTA 114

## RESULT 21

AAI12064/c

ID AAI12064 standard; DNA; 429 BP.

XX AC AAI12064;

XX DT 12-OCT-2001 (first entry)

XX DE Probe #1997 for gene expression analysis in human cervical cell sample.

XX KW Probe; human; microarray; gene expression; cervical epithelial cell;

XX KW cervical cancer; ss.

XX OS Homo sapiens.

XX PN WO200157278-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00670.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488901/53.

XX PT Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human cervical epithelial cells -

XX PS Claim 25; SEQ ID No 1997; 487pp; English.

XX CC The present invention relates to human single exon nucleic acid probes  
XX (SENPs). The present sequence is one such probe. The SENPs are derived  
XX from human HeLa cells. The SENPs can be used to produce a single exon  
XX microarray, which can be used for measuring human gene expression in a  
XX sample derived from human cervical epithelial cells. By measuring gene  
XX expression, the probes are therefore useful in grading and/or staging  
XX of diseases of the cervix, notably cervical cancer.

XX CC Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 429 BP; 148 A; 78 C; 69 G; 134 T; 0 other;

## Query Match

Best Local Similarity 2.1%; Score 18; DB 22; Length 429;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 635 AAGGAAAAATAGGTATTA 652

Db 131 AAGGAAAAATAGGTATTA 114

## RESULT 22

AAI33398/c

ID AAI33398 standard; DNA; 429 BP.

XX AC AAI33398;

XX DT 17-OCT-2001 (first entry)

XX DE Probe #2084 used to measure gene expression in human placenta sample.

XX KW Probe; microarray; human; placenta; antenatal diagnosis;

```
KW genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US00663.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX
XX PR 26-MAY-2000; 2000US-0207456.
XX
XX PR 30-JUN-2000; 2000US-0608408.
XX
XX PR 03-AUG-2000; 2000US-0632366.
XX
XX PR 21-SEP-2000; 2000US-0234687.
XX
XX PR 27-SEP-2000; 2000US-0236359.
XX
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX DR WPI; 2001-488897/53.
XX
XX PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX PS Claim 25; SEQ ID No 2084; 654bp; English.
XX
XX CC The present invention relates to single exon nucleic acid probes (SENP).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
XX SQ Sequence 429 BP; 148 A; 78 C; 69 G; 134 T; 0 other;

Query Match 2.1%; Score 18; DB 22; Length 429;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 AAGGAAAATAGGTATTA 652
Db 131 AAGGAAAATAGGTATTA 114

RESULT 23
AAI01986/c
ID AAI01986 standard; DNA; 429 BP.
XX
XX AC AAI01986;
XX
XX DT 09-OCT-2001 (first entry)
XX
XX DE Probe #1977 used to measure gene expression in human breast sample.
XX
XX KW Probe; human; breast disease; breast cancer; development disorder; ss;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX OS Homo sapiens.
XX
XX PN WO200157270-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 29-JAN-2001; 2001WO-US00661.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX
XX PR 26-MAY-2000; 2000US-0207456.
XX
XX PR 30-JUN-2000; 2000US-0608408.
XX
XX PR 03-AUG-2000; 2000US-0632366.
XX
XX PR 21-SEP-2000; 2000US-0234687.
XX
XX PR 27-SEP-2000; 2000US-0236359.
XX
XX PR 04-OCT-2000; 2000US-0236359.
XX
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX DR WPI; 2001-476286/51.
XX
XX PT Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast -
XX
XX PS Claim 25; SEQ ID No 1977; 322pp; English.
XX
XX CC The present invention relates to novel single exon nucleic acid probes.
XX The present sequence is one such probe. The probes are useful for
XX measuring human gene expression in a human breast sample, where the probe
XX hybridises at high stringency to a nucleic acid expressed in the human
XX breast. The probes are useful for predicting, diagnosing, grading,
XX staging, monitoring and prognosing diseases of the human breast,
XX particularly those diseases with polygenic aetiology. The diseases
XX include: breast cancer, disorders of development, inflammatory diseases
XX of the breast, fibrocystic changes, proliferative breast disease and
XX non-carcinoma tumours.
XX
XX CC Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 429 BP; 148 A; 78 C; 69 G; 134 T; 0 other;

Query Match 2.1%; Score 18; DB 22; Length 429;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 AAGGAAAATAGGTATTA 652
Db 131 AAGGAAAATAGGTATTA 114

RESULT 24
ABS01971/c
ID ABS01971 standard; DNA; 429 BP.
XX
XX AC ABS01971;
XX
XX DT 19-AUG-2002 (first entry)
XX
XX DE Human genome-derived single exon probe from lung SEQ ID No 1962.
XX
XX KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease.
XX
XX OS Homo sapiens.
XX
XX PN WO200186003-A2.
XX
XX PD 15-NOV-2001.
XX
XX PF 30-JAN-2001; 2001WO-US00665.
XX
XX PR 04-FEB-2000; 2000US-180312P.
XX
XX PR 26-MAY-2000; 2000US-207456P.
XX
XX PR 30-JUN-2000; 2000US-0608408.
XX
XX PR 03-AUG-2000; 2000US-0632366.
XX
XX PR 21-SEP-2000; 2000US-234687P.
XX
XX PR 27-SEP-2000; 2000US-236359P.
XX
XX PR 04-OCT-2000; 2000GB-0024263.
```

XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX DR WPI; 2002-114183/15.  
XX PT Spatially-addressable set of single exon nucleic acid probes, used to  
XX PT measure gene expression in human lung samples -  
XX PS Claim 1; SEQ ID No 1962; 634pp; English.  
XX CC The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of  
CC probes; the novel set of probes which hybridise at high stringency to a  
CC nucleic acid expressed in the human lung; measuring gene expression in a  
CC sample derived from human lung, comprising (a) contacting the array with  
CC a collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of  
CC the array; identifying exons in a eukaryotic genome, comprising  
CC (a) algorithmically predicting at least one exon from genomic sequences  
CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene  
CC expression analysis, and for identifying exons in a gene, particularly  
CC using human lung derived mRNA and for the study of lung diseases  
CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
CC pulmonary alveolar proteinosis, Karagenen syndrome, fibrocystic  
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
CC and hyaline membrane disease. The present sequence is a single exon  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 429 BP; 148 A; 78 C; 69 G; 134 T; 0 other;  
  
Query Match 2.1%; Score 18; DB 24; Length 429;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 635 AAGGAAATAGGTATTA 652  
Db 131 AAGGAAATAGGTATTA 114  
  
RESULT 25  
AAC36822  
ID AAC36822 standard; DNA; 493 BP.  
XX AC AAC36822;  
XX AC AAC36822;  
DT 17-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 15193.

XX KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
XX metabolic pathway; promoter; termination sequence; ss.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126284.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.

```
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.

PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.1%; Score 18; DB 21; Length 493;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 808 TTGCGCGAGAAATTGGA 825
    |||||
Db 324 TTGCGCGAGAAATTGGA 341

RESULT 26
ABQ31518/c
ID ABQ31518 standard; DNA; 541 BP.
XX
AC ABQ31518;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 18109.
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
```



```

XX WO200218632-A2.
XX 07-MAR-2002.
XX 01-SEP-2001; 2001WO-EP10074.
XX 01-SEP-2000; 2000DE-1043826.
XX 05-SEP-2000; 2000DE-1044543.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX Determining the degree of cytosine methylation in genomic DNA, useful
XX for diagnosis and prognosis, comprises selective hybridization of
XX amplicons from chemically treated DNA
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one
XX member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
XX and the degree of hybridisation to both classes is determined from the
XX label on the amplicon. From the ratio of labels hybridised to the two
XX classes of oligomers, the degree of methylation is calculated. The method
XX is used: (i) for diagnosis and/or prognosis of side effects of
XX therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
XX of the central nervous, cardiovascular, gastrointestinal and respiratory
XX systems etc., particularly by detecting mutations or single nucleotide
XX polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
XX types and for investigating cell differentiation. The method allows the
XX methylation status of many C residues to be determined simultaneously.
XX ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
XX method for determining the degree of cytosine methylation described in
XX the disclosure of the invention.
XX Sequence 541 BP; 99 A; 52 C; 169 G; 221 T; 0 other;
XX Query Match 2.1%; Score 18; DB 24; Length 541;
XX Best Local Similarity 100.0%; Pred. No. 93;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 375 ATTTAATCCAAAAAACAC 392
Db 523 ATTTAATCCAAAAAACAC 506
RESULT 27
ABQ31519
ID ABQ31519 standard; DNA; 541 BP.
XX AC
XX ABQ31519;
XX 12-JUL-2002 (first entry)
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 18110.
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX Homo sapiens.
XX WO200218632-A2..

```

```

PD 07-MAR-2002.
XX 01-SEP-2001; 2001WO-EP10074.
XX 01-SEP-2000; 2000DE-1043826.
XX 05-SEP-2000; 2000DE-1044543.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX Determining the degree of cytosine methylation in genomic DNA, useful
XX for diagnosis and prognosis, comprises selective hybridization of
XX amplicons from chemically treated DNA
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one
XX member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
XX and the degree of hybridisation to both classes is determined from the
XX label on the amplicon. From the ratio of labels hybridised to the two
XX classes of oligomers, the degree of methylation is calculated. The method
XX is used: (i) for diagnosis and/or prognosis of side effects of
XX therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
XX of the central nervous, cardiovascular, gastrointestinal and respiratory
XX systems etc., particularly by detecting mutations or single nucleotide
XX polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
XX types and for investigating cell differentiation. The method allows the
XX methylation status of many C residues to be determined simultaneously.
XX ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
XX method for determining the degree of cytosine methylation described in
XX the disclosure of the invention.
XX Sequence 541 BP; 221 A; 169 C; 52 G; 99 T; 0 other;
XX Query Match 2.1%; Score 18; DB 24; Length 541;
XX Best Local Similarity 100.0%; Pred. No. 93;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 375 ATTTAATCCAAAAAACAC 392
Db 19 ATTTAATCCAAAAAACAC 36
RESULT 28
AAZ22845
ID AAZ22845 standard; DNA; 586 BP.
XX AC
XX AAZ22845;
XX 20-DEC-1999 (first entry)
XX Staphylococcus aureus ferric uptake regulator B (furB) DNA.
XX Infection; detection; diagnosis; screening; antibiotic; resistance;
XX methicillin; MRSA; ds.
XX Staphylococcus aureus.
XX Key Location/Qualifiers
XX CDS 101..547
XX /*tag= a
XX /product= "S. aureus ferric uptake regulator B (furB)
XX protein"
XX WO9947662-A1.

```

```

XX 23-SEP-1999.
XX PD WO9947639-A2.
XX PN
XX PD 23-SEP-1999.
XX PF
XX PD 19-MAR-1999; 99WO-US05976.
XX PR
XX PR 20-MAR-1998; 98US-0078682.
XX PR 01-APR-1998; 98US-0080296.
XX PR 07-MAY-1998; 98US-0084674.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PI Simpson AJG, Choi GH;
XX WPI; 1999-580304/49.
XX DR P-PSDB; AAY42326.
XX STaphylococcus aureus genes and polypeptides, vectors and methods of
XX recombinant production
XX PS Claim 1; Page 19; 11pp; English.
XX CC This sequence represents Staphylococcus aureus ferric uptake regulator
XX B (furb) DNA. The furb protein is predicted to have a molecular
XX weight of about 17.2 kD and, along with furA (AAY42325) and furC
XX (AAY42327), is thought to be involved in iron regulation, based on amino
XX acid sequence homology with known ferric uptake regulators in proteins. S.
XX aureus is a ubiquitous pathogen which causes infections in burns,
XX cellulitis, eyelid infections, food poisoning, joint infections,
XX neonatal conjunctivitis, osteomyelitis, skin infections, surgical wound
XX infection, scalded skin syndrome and toxic shock syndrome. S. aureus is
XX increasingly becoming resistant to known antibiotics, with
XX methicillin-resistant strains generally being multiply drug resistant.
XX Methicillin-resistant S. aureus (MRSA) poses serious infection control
XX problems, with many strains being multiresistant against virtually all
XX antibiotics. With the exception of the vancomycin-type glycopeptide
XX antibiotics. The protein may be useful to screen potential antagonists
XX which could be used as antibiotics and it may be used as a vaccine to
XX prevent or attenuate an infection caused by a member of the
XX Staphylococcus genus. The protein, or antibodies against it
XX can be used in immunoassays to detect Staphylococcus in a biological
XX sample. Probes and primers derived from the nucleic acid sequences may
XX also be used to detect Staphylococcus nucleic acids in a biological
XX sample.
XX SQ Sequence 586 BP; 232 A; 78 C; 92 G; 184 T; 0 other;

Query Match 2.1%; Score 18; DB 20; Length 586;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 AAGAAAGATGGTGATATA 243
Db 82 AAGAAAGATGGTGATATA 99

RESULT 29
AAZ19884
ID AAZ19884 standard; DNA; 586 BP.
XX AC
XX AAZ19884;
XX 06-DEC-1999 (first entry)
XX STaphylococcus aureus ferric uptake regulator furb gene.
XX Furb gene; infection; therapy; diagnosis; vaccine; antibiotic;
XX ferric uptake regulator; ss.
XX STaphylococcus aureus.
XX Key Location/Qualifiers
XX CDS 101..547

FT
XX
XX WO9947639-A2.
XX PN
XX PD 23-SEP-1999.
XX PF
XX PD 19-MAR-1999; 99WO-US05976.
XX PR
XX PR 20-MAR-1998; 98US-0078682.
XX PR 01-APR-1998; 98US-0080296.
XX PR 07-MAY-1998; 98US-0084674.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PI Simpson AJG, Choi GH;
XX WPI; 1999-562101/47.
XX DR P-PSDB; AAY31819.
XX STaphylococcus aureus nucleic acid molecules, used to develop
XX products for the diagnosis, prevention and treatment of Staphylococcal
XX infections
XX PS Claim 1; Page 19; 102pp; English.
XX CC This is the nucleotide sequence of Staphylococcus aureus strain
XX ISP3 (ATCC 202108) genomic DNA including the novel furb gene that
XX codes for a 148-amino acid protein (see AAY31819) of predicted
XX mol. wt. 17.2 kDa. The sequence was obtained from isolated DNA
XX clone BTALE70. Furb is a ferric uptake regulator that shows
XX amino acid sequence homology to known proteins involved in iron
XX regulation. The invention provides 11 novel genes (see AAZ19882-92)
XX of S. aureus and the polypeptides they encode (see AAY31817-27). Also
XX provided are vectors, host cells, antibodies and hybridomas. The
XX invention further relates to screening methods for identifying
XX agonists and antagonists of S. aureus polypeptide activity, and to
XX diagnostic methods for detecting Staphylococcus nucleic acids, and to
XX polypeptides and antibodies in a biological sample. Antagonists
XX of furb may be useful as antibiotics to treat infections of
XX S. aureus and other Staphylococcus spp. Also provided are novel
XX vaccines for the prevention or attenuation of infection by
XX Staphylococcus. The isolated nucleic acid molecules are also
XX useful for generating probes and primers, and in the recombinant
XX production of furb protein.
XX SQ Sequence 586 BP; 232 A; 78 C; 92 G; 184 T; 0 other;

Query Match 2.1%; Score 18; DB 20; Length 586;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 AAGAAAGATGGTGATATA 243
Db 82 AAGAAAGATGGTGATATA 99

RESULT 30
ABA63236
ID ABA63236 standard; DNA; 588 BP.
XX AC
XX ABA63236;
XX 01-FEB-2002 (first entry)
XX Human foetal liver single exon nucleic acid probe #11541.
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX Homo sapiens.
XX WO200157277-A2.
XX

```

```
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
XX Claim 1; SEQ ID NO 11541; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 588 BP; 210 A; 112 C; 120 G; 146 T; 0 other;
SQ
Query Match 2.1%; Score 18; DB 22; Length 588;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 461 AAGATCAGCAATATGTAG 478
Db 357 AAGATCAGCAATATGTAG 374
Query Match 2.1%; Score 18; DB 22; Length 588;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 461 AAGATCAGCAATATGTAG 478
Db 357 AAGATCAGCAATATGTAG 374
RESULT 31
ABA30473
ID ABA30473 standard; DNA; 588 BP.
XX
XX ABA30473;
AC
XX
XX 23-JAN-2002 (first entry)
DT
XX
XX Probe #8939 for gene expression analysis in human heart cell sample.
DE
XX
XX Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200157274-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00666.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
XX Claim 1; SEQ ID NO 11541; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 588 BP; 210 A; 112 C; 120 G; 146 T; 0 other;
SQ
Query Match 2.1%; Score 18; DB 22; Length 588;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 461 AAGATCAGCAATATGTAG 478
Db 357 AAGATCAGCAATATGTAG 374
RESULT 32
AAK11702
ID AAK11702 standard; DNA; 588 BP.
XX
XX AAK11702;
AC
XX
XX 05-NOV-2001 (first entry)
DT
XX
XX Human brain expressed single exon probe SEQ ID NO: 11693.
DE
XX
XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200157275-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00667.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT
```

```
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
XX Claim 1; SEQ ID No 8939; 530pp; English.
PS
XX
XX The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart, and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 588 BP; 210 A; 112 C; 120 G; 146 T; 0 other;
SQ
Query Match 2.1%; Score 18; DB 22; Length 588;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 461 AAGATCAGCAATATGTAG 478
Db 357 AAGATCAGCAATATGTAG 374
RESULT 32
AAK11702
ID AAK11702 standard; DNA; 588 BP.
XX
XX AAK11702;
AC
XX
XX 05-NOV-2001 (first entry)
DT
XX
XX Human brain expressed single exon probe SEQ ID NO: 11693.
DE
XX
XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200157275-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00667.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT
```

```

PT brains -
XX Example 4; SEQ ID NO: 11693; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 588 BP; 210 A; 112 C; 120 G; 146 T; 0 other;

Query Match      2.1%; Score 18; DB 22; Length 588;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 461 AAGATCAGCAATATGTAG 478
DB 357 AAGATCAGCAATATGTAG 374

RESULT 33
AAK37443
ID AAK37443 standard; DNA; 588 BP.
XX
AC AAK37443;
XX
DT 06-NOV-2001 (first entry)
DE Human bone marrow expressed single exon probe SEQ ID NO: 12000.
XX
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Claim 25; SEQ ID NO 8190; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 588 BP; 210 A; 112 C; 120 G; 146 T; 0 other;

Query Match      2.1%; Score 18; DB 22; Length 588;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 461 AAGATCAGCAATATGTAG 478
DB 357 AAGATCAGCAATATGTAG 374

RESULT 35
AAI43306
ID AAI43306 standard; DNA; 588 BP.
XX

```

```

Best Local Similarity 100.0%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 461 AAGATCAGCAATATGTAG 478
DB 357 AAGATCAGCAATATGTAG 374

RESULT 34
AAI18257
ID AAI18257 standard; DNA; 588 BP.
XX
AC AAI18257;
XX
DT 12-OCT-2001 (first entry)
DE Probe #8190 for gene expression analysis in human cervical cell sample.
XX
XX Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID NO 8190; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 588 BP; 210 A; 112 C; 120 G; 146 T; 0 other;

Query Match      2.1%; Score 18; DB 22; Length 588;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 461 AAGATCAGCAATATGTAG 478
DB 357 AAGATCAGCAATATGTAG 374

RESULT 35
AAI43306
ID AAI43306 standard; DNA; 588 BP.
XX

```

```
XX AC AAI43306;
XX DT 17-OCT-2001 (first entry)
XX DE Probe #11992 used to measure gene expression in human placenta sample.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488897/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta -
XX PS Claim 25; SEQ ID No 11992; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENPs).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders.
XX SQ Sequence 588 BP; 210 A; 112 C; 120 G; 146 T; 0 other;

Query Match          2.1%; Score 18; DB 22; Length 588;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 461 AAGATCAGCAATATGTAG 478
Db 357 AAGATCAGCAATATGTAG 374

RESULT 36
ABS11430
ID ABS11430 standard; DNA; 588 BP.
XX AC ABS11430;
XX AC ABS11430;
XX DT 19-AUG-2002 (first entry)
XX DE Human genome-derived single exon probe from lung SEQ ID No 11421.
XX KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
XX KW chronic obstructive pulmonary disease; interstitial lung disease;
XX KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
XX KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX KW primary ciliary dyskinesia; pulmonary hypertension;
XX KW hyaline membrane disease.
```

```
XX OS Homo sapiens.
XX PN WO200186003-A2.
XX PD 15-NOV-2001.
XX PF 30-JAN-2001; 2001WO-US00665.
XX PR 04-FEB-2000; 2000US-180312P.
XX PR 26-MAY-2000; 2000US-207456P.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-234687P.
XX PR 27-SEP-2000; 2000US-236359P.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2002-114183/15.
XX PT Spatially-addressable set of single exon nucleic acid probes, used to
XX PT measure gene expression in human lung samples -
XX PS Claim 1; SEQ ID No 11421; 634pp; English.
XX CC The invention relates to a spatially-addressable set of single exon
XX CC nucleic acid probes for measuring gene expression in a sample derived
XX CC from human lung comprising single exon nucleic acid probes having one of
XX CC 12614 nucleic acid sequences mentioned in the specification, or their
XX CC complements or the 12387 open reading frames derived from the 12614
XX CC probes. Also included are a microarray comprising the novel set of
XX CC probes; the novel set of probes which hybridise at high stringency to a
XX CC nucleic acid expressed in the human lung; measuring gene expression in a
XX CC sample derived from human lung, comprising (a) contacting the array with
XX CC a collection of detectably labeled nucleic acids derived from human lung
XX CC mRNA, and (b) measuring the label detectably bound to each probe of
XX CC the array; identifying exons in a eukaryotic genome, comprising
XX CC (a) algorithmically predicting at least one exon from genomic sequences
XX CC of the eukaryote; and (b) detecting specific hybridisation of detectably
XX CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
XX CC having a fragment identical to the predicted exon, the probe is included
XX CC in the above mentioned microarray; assigning exons to a single gene,
XX CC comprising (a) identifying exons from genomic sequence by the method
XX CC above and (b) measuring the expression of each of the exons in several
XX CC tissues and/or cell types using hybridisation to a single exon
XX CC microarrays having a probe with the exon, where a common pattern of
XX CC expression of the exons in the tissues and/or cell types indicates that
XX CC the exons should be assigned to a single gene; a peptide comprising one
XX CC of 12011 sequences, mentioned in the specification, or encoded by the
XX CC probes/open reading frames (ORF). The probes are used for gene
XX CC expression analysis, and for identifying exons in a gene, particularly
XX CC using human lung derived mRNA and for the study of lung diseases
XX CC such as asthma, lung cancer, chronic obstructive pulmonary disease
XX CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
XX CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
XX CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
XX CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
XX CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
XX CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
XX CC and hyaline membrane disease. The present sequence is a single exon
XX CC probe of the invention.
XX CC Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic
XX CC format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 588 BP; 210 A; 112 C; 120 G; 146 T; 0 other;

Query Match          2.1%; Score 18; DB 24; Length 588;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 461 AAGATCAGCAATATAG 478
DB 357 AAGATCAGCAATATAG 374

RESULT 37
ABV50953/c
ID ABV50953 standard; cDNA; 601 BP.
XX
AC ABV50953;
XX
DT 17-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 50944.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
XX
PR 16-MAR-2000; 2000US-189862P.
XX
PR 25-MAY-2000; 2000US-207454P.
XX
PR 09-JUN-2000; 2000US-211314P.
XX
PR 18-JUL-2000; 2000US-219007P.
XX
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX
Claim 1; Page 9907; 11750pp; English.
XX
PS The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
Sequence 601 BP; 210 A; 105 C; 139 G; 147 T; 0 other;

Query Match 2.1%; Score 18; DB 23; Length 601;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 TATAAACATTTTGCAATTA 441
DB 523 TATAAACATTTTGCAATTA 506

RESULT 38
AAX03394
ID AAX03394 standard; cDNA; 800 BP.
XX
AC AAX03394;
XX
DT 31-MAR-1999 (first entry)
XX
DE Nucleotide sequence of the 3' terminus of MCDV-M1.
XX
KW Maize chlorotic dwarf virus strain Tennessee; MCDV-Tn; MCDV-T; MCDV-M1;
KW coat protein; transgenic plant; resistance; MCDV-Tn infection; ss.
XX
OS Maize chlorotic dwarf virus.
XX
PN US5866780-A.
XX
PD 02-FEB-1999.
XX
PF 04-APR-1995; 95US-0416603.
XX
PR 04-APR-1995; 95US-0416603.
XX
PA (NOVS) NOVARTIS FINANCE CORP.
XX
PI Habera L, Law M, Reddick BB;
XX
DR WPI; 1999-142034/12.
XX
PT cDNA of maize chlorotic dwarf virus strain MCDV-Tn - and chimeric
PT genes for producing e.g. MCDV-Tn-resistant transgenic plants
XX
PS Disclosure; Columns 143-144; 73pp; English.
XX
CC The present sequence represents the nucleotide sequence of the 3'
CC terminus of the maize chlorotic dwarf virus strain M1 (MCDV-M1). The
CC specification describes a MCDV strain that is distinct from known MCDV-T
CC and MCDV-M1 isolates, called MCDV Tennessee (MCDV-Tn). The specification
CC also provides the sequences for 3 MCDV-Tn coat proteins, which may
CC be expressed in transgenic plants to confer resistance to MCDV-Tn
CC infection. The coat proteins can also be produced recombinantly.
XX
SQ Sequence 800 BP; 196 A; 153 C; 215 G; 236 T; 0 other;

Query Match 2.1%; Score 18; DB 20; Length 800;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 AGTAGCTCCAGGCATTGA 284
DB 394 AGTAGCTCCAGGCATTGA 411

RESULT 39
AAC42971
ID AAC42971 standard; DNA; 933 BP.
XX
AC AAC42971;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 37548.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
```

XX 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134766.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140354.  
PR 23-JUN-1999; 99US-0140355.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145132.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147036.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151085.  
PR 27-AUG-1999; 99US-0151086.  
PR 27-AUG-1999; 99US-0151087.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.

PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 2.1%; Score 18; DB 21; Length 933;  
 Best Local Similarity 100.0%; Pred. No. 92;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 808 TTTCGGCGAGAAATTGGA 825  
 |||||  
 DB 249 TTTCGGCGAGAAATTGGA 266

## RESULT 40

AAT95372  
 ID AAT95372 standard; cDNA; 956 BP.

XX AC AAT95372;

XX DT 14-APR-1998 (first entry)

XX DE Arabidopsis thaliana beta-carotene hydroxylase cDNA.

XX KW Beta-carotene hydroxylase; carotenoid; biosynthetic enzyme;  
 pigment; vector; pATOHB; ss.

XX OS Arabidopsis thaliana.

XX FH Key Location/Qualifiers  
 FT CDS 3..887  
 FT /\*tag= a

XX PN WO9736998-A1.

XX PD 09-OCT-1997.

XX PF 28-JAN-1997; 97WO-US00540.

XX PR 29-MAR-1996; 96US-0624125.

XX (UYMA-) UNIV MARYLAND BALTIMORE.  
 PA Cunningham FX, Sun Z;  
 PI WPI; 1997-503091/46.  
 DR P-PSDB; AAW35347.  
 XX Eukaryotic carotenoid biosynthetic enzymes and related genes -  
 PT useful to control ratio of various carotenoid(s) in host and for  
 PT production of novel carotenoid pigments  
 XX Claim 17; Page 37-38; 89pp; English.  
 XX This cDNA sequence codes for the beta-carotene hydroxylase (see  
 CC AAW35347) of Arabidopsis thaliana, an enzyme responsible for  
 CC hydroxylating the beta end-group in carotenoids. It was isolated  
 CC from an Arabidopsis cDNA expression library through zeaxanthin  
 CC pigment production. Expression vector pATOHB comprising the  
 CC beta-carotene hydroxylase gene is deposited as ATCC 98003. Claimed  
 CC isolated DNA sequences (see AAT95371-73 and AAT95389-91) encoding the  
 CC claimed eukaryotic carotenoid biosynthetic enzymes epsilon cyclase,  
 CC beta-carotene hydroxylase and isopentenyl pyrophosphate isomerase  
 CC (see AAW35346-51) are used in methods for augmenting the accumulation  
 CC of carotenoids and for the production of novel and rare carotenoids  
 CC in host cells. Methods are also provided for controlling the ratio  
 CC of various carotenoids in a host, and for screening for eukaryotic  
 CC genes that encode enzymes of carotenoid biosynthesis and metabolism.  
 XX SQ Sequence 956 BP; 224 A; 216 C; 234 G; 282 T; 0 other;

Query Match 2.1%; Score 18; DB 18; Length 956;  
 Best Local Similarity 100.0%; Pred. No. 92;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 808 TTTCGGCGAGAAATTGGA 825  
 |||||  
 DB 203 TTTCGGCGAGAAATTGGA 220

Search completed: July 4, 2003, 01:47:17  
 Job time : 227 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 01:41:02 ; Search time 46 Seconds  
(without alignments)  
5600.184 Million cell updates/sec

Title: US-10-062-624-41

Perfect score: 840

Sequence: 1 atgaattataagaataattctt.....ttggaatgaggttcaccttc 840

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued Patents NA.\*

- 1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	840	100.0	840	4	US-09-660-587-41
2	22	2.6	22	4	US-09-660-587-37
3	19	2.3	840	4	US-09-660-587-5
4	19	2.3	840	4	US-09-261-358A-5
5	18	2.1	800	2	US-08-416-603-11
6	18	2.1	956	1	US-08-624-125-3
7	18	2.1	4673	1	US-07-638-431-1
8	18	2.1	4673	5	PCT-US92-00018-1
9	18	2.1	152331	3	US-09-128-155-16
10	18	2.1	176373	3	US-09-128-155-17
11	17	2.0	24	3	US-08-755-587-151
12	17	2.0	75	3	US-09-023-082A-35
13	17	2.0	90	3	US-09-023-082A-103
14	17	2.0	90	4	US-09-218-444-24
15	17	2.0	177	4	US-09-581-617-4
16	17	2.0	276	3	US-09-023-082A-81
17	17	2.0	357	3	US-09-023-082A-79
18	17	2.0	426	3	US-09-023-082A-69
19	17	2.0	426	3	US-09-023-082A-95
20	17	2.0	426	3	US-09-023-082A-111
21	17	2.0	426	3	US-09-023-082A-119
22	17	2.0	426	3	US-09-023-082A-123
23	17	2.0	426	3	US-09-023-082A-127
24	17	2.0	426	3	US-09-023-082A-131
25	17	2.0	426	3	US-09-023-082A-135
26	17	2.0	426	3	US-09-023-082A-139
27	17	2.0	426	3	US-09-023-082A-145

Sequence 16, Appl  
Sequence 32, Appl  
Sequence 67, Appl  
Sequence 115, Appl  
Sequence 42, Appl  
Sequence 65, Appl  
Sequence 83, Appl  
Sequence 84, Appl  
Sequence 28, Appl  
Sequence 38, Appl  
Sequence 40, Appl  
Sequence 39, Appl  
Sequence 38, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 47, Appl  
Sequence 1, Appl  
Sequence 63, Appl  
Sequence 64, Appl  
Sequence 63, Appl  
Sequence 64, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 105, Appl  
Sequence 50, Appl  
Sequence 7, Appl  
Sequence 11, Appl  
Sequence 7, Appl  
Sequence 11, Appl  
Sequence 56, Appl  
Sequence 32, Appl  
Sequence 2, Appl  
Sequence 20, Appl  
Sequence 20, Appl  
Sequence 2078, Appl  
Sequence 105, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 7, Appl  
Sequence 7, Appl  
Sequence 9, Appl  
Sequence 9, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 3, Appl  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 5, Appl  
Sequence 5, Appl  
Sequence 4, Appl  
Sequence 12, Appl  
Patent No. 5194375  
Sequence 2, Appl  
Sequence 4, Appl  
Sequence 2, Appl  
Sequence 1, Appl  
Sequence 1, Appl

C 101	16	1.9	3115	4	US-09-215-649A-1	Sequence 1, Appl	174	15	1.8	107	3	US-08-258-287B-23	Sequence 23, Appl
C 102	16	1.9	3115	4	US-09-577-780-1	Sequence 1, Appl	175	15	1.8	107	3	US-08-388-704C-23	Sequence 23, Appl
C 103	16	1.9	3136	3	US-08-996-129-3	Sequence 5, Appl	176	15	1.8	183	4	US-09-134-001C-2550	Sequence 2550, Ap
C 104	16	1.9	3136	4	US-09-435-139-5	Sequence 3, Appl	C 177	15	1.8	212	1	US-08-392-678-18	Sequence 18, Appl
C 105	16	1.9	3136	4	US-08-995-653-5	Sequence 5, Appl	C 178	15	1.8	212	1	US-08-392-678-28	Sequence 28, Appl
C 106	16	1.9	3136	4	US-09-215-649A-5	Sequence 5, Appl	C 179	15	1.8	212	1	US-08-457-304A-18	Sequence 28, Appl
C 107	16	1.9	3136	4	US-09-577-780-5	Sequence 5, Appl	C 180	15	1.8	212	1	US-08-457-304A-28	Sequence 28, Appl
C 108	16	1.9	3340	4	US-09-021-560-1	Sequence 1, Appl	C 181	15	1.8	212	1	US-08-456-701A-18	Sequence 28, Appl
C 109	16	1.9	3469	4	US-09-221-017B-106	Sequence 106, App	C 182	15	1.8	212	1	US-08-456-701A-28	Sequence 28, Appl
C 110	16	1.9	3329	4	US-09-508-542-16	Sequence 16, Appl	C 183	15	1.8	212	4	US-08-684-932A-18	Sequence 18, Appl
C 111	16	1.9	3620	4	US-09-446-504-55	Sequence 55, Appl	C 184	15	1.8	256	3	US-08-616-368A-25	Sequence 25, Appl
C 112	16	1.9	3620	4	US-09-712-266-55	Sequence 55, Appl	C 185	15	1.8	256	3	US-09-054-298-25	Sequence 25, Appl
C 113	16	1.9	4334	2	US-08-670-707A-38	Sequence 38, Appl	C 186	15	1.8	256	4	US-08-818-655-25	Sequence 25, Appl
C 114	16	1.9	4334	4	US-09-037-601-38	Sequence 38, Appl	C 187	15	1.8	277	3	US-08-746-111-54	Sequence 54, Appl
C 115	16	1.9	4334	4	US-09-315-179-38	Sequence 38, Appl	C 188	15	1.8	299	4	US-09-328-111-349	Sequence 349, App
C 116	16	1.9	4404	4	US-09-523-656-37	Sequence 37, Appl	C 189	15	1.8	333	2	US-08-611-757-39	Sequence 39, Appl
C 117	16	1.9	4629	2	US-08-484-891-7	Sequence 7, Appl	C 190	15	1.8	333	5	PCT-US95-05980-39	Sequence 39, Appl
C 118	16	1.9	4670	3	US-08-717-294-41	Sequence 41, Appl	C 191	15	1.8	378	3	US-08-535-057A-5	Sequence 5, Appl
C 119	16	1.9	4999	4	US-09-470-618-14	Sequence 14, Appl	C 192	15	1.8	380	4	US-09-118-554-31	Sequence 31, Appl
C 120	16	1.9	4999	4	US-09-364-862-14	Sequence 14, Appl	C 193	15	1.8	380	4	US-09-118-627-31	Sequence 31, Appl
C 121	16	1.9	5035	2	US-08-882-083-1	Sequence 1, Appl	C 194	15	1.8	380	4	US-09-602-877A-31	Sequence 31, Appl
C 122	16	1.9	5035	2	US-08-558-107-1	Sequence 1, Appl	C 195	15	1.8	381	3	US-08-535-057A-4	Sequence 4, Appl
C 123	16	1.9	5035	3	US-09-243-536-1	Sequence 1, Appl	C 196	15	1.8	411	2	US-08-449-287-5	Sequence 5, Appl
C 124	16	1.9	5058	4	US-08-934-386-8	Sequence 8, Appl	C 197	15	1.8	411	2	US-08-449-287-7	Sequence 7, Appl
C 125	16	1.9	5892	3	US-08-755-587-27	Sequence 27, Appl	C 198	15	1.8	424	4	US-09-393-569-3	Sequence 3, Appl
C 126	16	1.9	6402	2	US-08-670-707A-36	Sequence 36, Appl	C 199	15	1.8	453	4	US-08-134-001C-665	Sequence 665, App
C 127	16	1.9	6402	4	US-09-037-601-36	Sequence 36, Appl	C 200	15	1.8	465	3	US-08-535-057A-7	Sequence 7, Appl
C 128	16	1.9	6402	4	US-09-315-179-36	Sequence 36, Appl	C 201	15	1.8	525	3	US-09-023-082A-54	Sequence 54, Appl
C 129	16	1.9	6402	4	US-09-523-656-29	Sequence 29, Appl	C 202	15	1.8	560	4	US-09-347-803-13	Sequence 13, Appl
C 130	16	1.9	6433	4	US-08-927-219-128	Sequence 128, App	C 203	15	1.8	567	1	US-07-982-650D-7	Sequence 7, Appl
C 131	16	1.9	7032	4	US-09-324-867-1	Sequence 1, Appl	C 204	15	1.8	567	1	US-08-331-379-7	Sequence 7, Appl
C 132	16	1.9	7056	1	US-08-121-208-1	Sequence 1, Appl	C 205	15	1.8	567	2	US-08-757-046A-9	Sequence 9, Appl
C 133	16	1.9	7351	1	US-08-224-391-83	Sequence 83, Appl	C 206	15	1.8	567	2	US-08-757-046A-13	Sequence 13, Appl
C 134	16	1.9	7351	1	US-08-184-304-83	Sequence 83, Appl	C 207	15	1.8	567	3	US-09-447-208-9	Sequence 9, Appl
C 135	16	1.9	7351	2	US-08-184-009-127	Sequence 127, App	C 208	15	1.8	567	3	US-09-447-208-13	Sequence 13, Appl
C 136	16	1.9	7351	2	US-08-566-398-39	Sequence 39, Appl	C 209	15	1.8	567	3	US-09-135-988-9	Sequence 9, Appl
C 137	16	1.9	7351	2	US-08-458-358-127	Sequence 127, App	C 210	15	1.8	567	3	US-09-135-988-13	Sequence 13, Appl
C 138	16	1.9	7351	2	US-08-658-663-39	Sequence 39, Appl	C 211	15	1.8	567	4	US-03-277-716-9	Sequence 9, Appl
C 139	16	1.9	7351	4	US-08-796-101-3	Sequence 3, Appl	C 212	15	1.8	567	4	US-09-277-716-13	Sequence 13, Appl
C 140	16	1.9	7351	4	US-08-460-736-127	Sequence 127, App	C 213	15	1.8	567	4	US-08-597-274A-9	Sequence 9, Appl
C 141	16	1.9	7351	4	US-09-085-273-39	Sequence 39, Appl	C 214	15	1.8	567	4	US-08-597-274A-13	Sequence 13, Appl
C 142	16	1.9	8241	6	5171844-1	Patent No. 5171844	C 215	15	1.8	567	4	US-08-908-909-9	Sequence 9, Appl
C 143	16	1.9	8967	1	US-08-366-851A-1	Sequence 1, Appl	C 216	15	1.8	567	4	US-08-908-909-13	Sequence 13, Appl
C 144	16	1.9	9009	1	US-07-864-004B-3	Sequence 3, Appl	C 217	15	1.8	567	4	US-09-609-161B-9	Sequence 9, Appl
C 145	16	1.9	9009	1	US-08-251-937A-3	Sequence 3, Appl	C 218	15	1.8	567	4	US-09-609-161B-13	Sequence 13, Appl
C 146	16	1.9	9009	1	US-08-212-133A-1	Sequence 1, Appl	C 219	15	1.8	567	4	US-08-990-103-9	Sequence 9, Appl
C 147	16	1.9	9009	1	US-08-474-503-1	Sequence 1, Appl	C 220	15	1.8	567	4	US-08-990-103-13	Sequence 13, Appl
C 148	16	1.9	9009	2	US-08-670-707A-1	Sequence 1, Appl	C 221	15	1.8	588	1	US-07-982-650D-6	Sequence 6, Appl
C 149	16	1.9	9009	4	US-09-037-601-1	Sequence 1, Appl	C 222	15	1.8	588	1	US-07-982-650D-8	Sequence 8, Appl
C 150	16	1.9	9009	4	US-09-315-179-1	Sequence 1, Appl	C 223	15	1.8	588	1	US-07-982-650D-9	Sequence 9, Appl
C 151	16	1.9	9009	4	US-09-523-656-1	Sequence 1, Appl	C 224	15	1.8	588	1	US-07-982-650D-10	Sequence 10, Appl
C 152	16	1.9	9009	5	PCT-US93-03275-3	Sequence 3, Appl	C 225	15	1.8	588	1	US-08-331-379-6	Sequence 6, Appl
C 153	16	1.9	9009	5	PCT-US94-13200-1	Sequence 1, Appl	C 226	15	1.8	588	1	US-08-331-379-8	Sequence 8, Appl
C 154	16	1.9	9354	1	US-08-683-839B-2	Sequence 2, Appl	C 227	15	1.8	588	1	US-08-331-379-9	Sequence 9, Appl
C 155	16	1.9	9361	4	US-08-934-386-7	Sequence 7, Appl	C 228	15	1.8	588	1	US-08-331-379-10	Sequence 10, Appl
C 156	16	1.9	9432	1	US-08-277-233A-1	Sequence 1, Appl	C 229	15	1.8	588	2	US-08-757-046A-10	Sequence 10, Appl
C 157	16	1.9	9432	2	US-08-473-750-4	Sequence 4, Appl	C 230	15	1.8	588	2	US-08-757-046A-11	Sequence 11, Appl
C 158	16	1.9	9432	2	US-08-477-326-4	Sequence 4, Appl	C 231	15	1.8	588	2	US-08-757-046A-12	Sequence 12, Appl
C 159	16	1.9	10815	4	US-09-004-838-21	Sequence 21, Appl	C 232	15	1.8	588	3	US-09-447-208-10	Sequence 10, Appl
C 160	16	1.9	11832	2	US-08-416-603-1	Sequence 1, Appl	C 233	15	1.8	588	3	US-09-447-208-11	Sequence 11, Appl
C 161	16	1.9	11933	4	US-09-470-618-13	Sequence 13, Appl	C 234	15	1.8	588	3	US-09-447-208-12	Sequence 12, Appl
C 162	16	1.9	11933	4	US-09-364-862-13	Sequence 13, Appl	C 235	15	1.8	588	3	US-03-135-988-10	Sequence 10, Appl
C 163	16	1.9	13149	4	US-09-004-838-87	Sequence 87, Appl	C 236	15	1.8	588	3	US-09-135-988-11	Sequence 11, Appl
C 164	16	1.9	17425	4	US-09-511-625B-5	Sequence 5, Appl	C 237	15	1.8	588	3	US-09-135-988-12	Sequence 12, Appl
C 165	16	1.9	35060	3	US-08-814-095-7	Sequence 7, Appl	C 238	15	1.8	588	4	US-09-277-716-10	Sequence 10, Appl
C 166	16	1.9	72928	3	US-09-009-913-1	Sequence 1, Appl	C 239	15	1.8	588	4	US-09-277-716-11	Sequence 11, Appl
C 167	16	1.9	168575	4	US-09-426-290-1	Sequence 1, Appl	C 240	15	1.8	588	4	US-09-277-716-12	Sequence 12, Appl
C 168	16	1.9	4403765	4	US-09-103-840A-2	Sequence 2, Appl	C 241	15	1.8	588	4	US-08-597-274A-10	Sequence 10, Appl
C 169	16	1.9	4411529	4	US-09-103-840A-1	Sequence 1, Appl	C 242	15	1.8	588	4	US-08-597-274A-11	Sequence 11, Appl
C 170	15	1.8	47	3	US-09-023-082A-51	Sequence 51, Appl	C 243	15	1.8	588	4	US-08-597-274A-12	Sequence 12, Appl
C 171	15	1.8	48	3	US-09-023-082A-47	Sequence 47, Appl	C 244	15	1.8	588	4	US-08-908-909-10	Sequence 10, Appl
C 172	15	1.8	70	2	US-08-894-578-114	Sequence 114, App	C 245	15	1.8	588	4	US-08-908-909-11	Sequence 11, Appl
C 173	15	1.8	98	4	US-08-976-183A-43	Sequence 43, Appl	C 246	15	1.8	588	4	US-08-908-909-12	Sequence 12, Appl

C 247	15	1.8	588	4	US-09-609-161B-10	Sequence 10, Appl	320	15	1.8	1160	1	US-08-290-448A-77	Sequence 77, Appl
C 248	15	1.8	588	4	US-09-609-161B-11	Sequence 11, Appl	321	15	1.8	1160	1	US-08-290-448A-77	Sequence 77, Appl
C 249	15	1.8	588	4	US-09-609-161B-12	Sequence 12, Appl	322	15	1.8	1160	1	US-08-175-069A-77	Sequence 77, Appl
C 250	15	1.8	588	4	US-08-990-103-10	Sequence 10, Appl	323	15	1.8	1160	4	US-08-461-939B-77	Sequence 77, Appl
C 251	15	1.8	588	4	US-08-990-103-11	Sequence 11, Appl	324	15	1.8	1160	4	US-08-461-939B-77	Sequence 77, Appl
C 252	15	1.8	588	4	US-08-990-103-12	Sequence 12, Appl	325	15	1.8	1178	2	US-08-933-750C-88	Sequence 88, Appl
C 253	15	1.8	588	4	US-08-990-103-13	Sequence 13, Appl	326	15	1.8	1178	3	US-09-234-613-88	Sequence 88, Appl
C 254	15	1.8	591	2	US-08-757-046A-6	Sequence 6, Appl	327	15	1.8	1178	3	US-09-234-613-88	Sequence 88, Appl
C 255	15	1.8	591	2	US-08-757-046A-7	Sequence 7, Appl	328	15	1.8	1196	1	US-08-290-448A-71	Sequence 71, Appl
C 256	15	1.8	591	3	US-09-447-208-6	Sequence 6, Appl	329	15	1.8	1196	1	US-08-290-448A-71	Sequence 71, Appl
C 257	15	1.8	591	3	US-09-447-208-7	Sequence 7, Appl	330	15	1.8	1196	1	US-08-175-069A-71	Sequence 71, Appl
C 258	15	1.8	591	3	US-08-535-057A-6	Sequence 6, Appl	331	15	1.8	1196	4	US-08-461-939B-71	Sequence 71, Appl
C 259	15	1.8	591	3	US-09-135-988-6	Sequence 6, Appl	332	15	1.8	1196	4	US-08-461-939B-71	Sequence 71, Appl
C 260	15	1.8	591	3	US-08-135-988-7	Sequence 7, Appl	C 332	15	1.8	1245	4	US-09-386-875-1	Sequence 1, Appl
C 261	15	1.8	591	4	US-08-597-274A-6	Sequence 6, Appl	C 333	15	1.8	1254	4	US-09-134-001C-2169	Sequence 2169, Ap
C 262	15	1.8	591	4	US-08-597-274A-7	Sequence 7, Appl	334	15	1.8	1270	1	US-08-378-588-23	Sequence 23, Appl
C 263	15	1.8	591	4	US-08-908-909-6	Sequence 6, Appl	335	15	1.8	1270	1	US-08-811-094-23	Sequence 23, Appl
C 264	15	1.8	591	4	US-08-908-909-7	Sequence 7, Appl	336	15	1.8	1270	5	PCT-US94-11121-23	Sequence 23, Appl
C 265	15	1.8	591	4	US-09-609-161B-6	Sequence 6, Appl	C 337	15	1.8	1295	1	US-08-332-747-1	Sequence 1, Appl
C 266	15	1.8	591	4	US-09-609-161B-7	Sequence 7, Appl	338	15	1.8	1332	3	US-09-221-235-15	Sequence 15, Appl
C 267	15	1.8	591	4	US-08-990-103-6	Sequence 6, Appl	339	15	1.8	1332	3	US-09-221-235-15	Sequence 15, Appl
C 268	15	1.8	591	4	US-08-990-103-7	Sequence 7, Appl	340	15	1.8	1332	3	US-09-221-235-15	Sequence 15, Appl
C 269	15	1.8	594	4	US-08-858-207A-115	Sequence 115, App	341	15	1.8	1332	3	US-09-221-235-15	Sequence 15, Appl
C 270	15	1.8	598	2	US-08-852-807-18	Sequence 18, Appl	342	15	1.8	1332	3	US-09-221-235-15	Sequence 15, Appl
C 271	15	1.8	599	4	US-09-277-716-5	Sequence 7, Appl	343	15	1.8	1332	4	US-09-221-235-15	Sequence 15, Appl
C 272	15	1.8	601	4	US-08-277-716-7	Sequence 7, Appl	344	15	1.8	1332	4	US-09-221-235-15	Sequence 15, Appl
C 273	15	1.8	615	4	US-09-328-111-142	Sequence 142, App	345	15	1.8	1332	4	US-09-163-115-15	Sequence 15, Appl
C 274	15	1.8	615	4	US-09-385-982-30	Sequence 30, Appl	346	15	1.8	1332	4	US-09-221-235-15	Sequence 15, Appl
C 275	15	1.8	619	4	US-09-328-111-1415	Sequence 415, App	347	15	1.8	1332	4	US-09-593-553-15	Sequence 15, Appl
C 276	15	1.8	651	4	US-09-257-179-12	Sequence 12, Appl	348	15	1.8	1332	4	US-09-221-235-15	Sequence 15, Appl
C 277	15	1.8	653	3	US-08-535-057A-2	Sequence 2, Appl	349	15	1.8	1333	3	US-09-221-235-15	Sequence 15, Appl
C 278	15	1.8	711	1	US-08-307-439-49	Sequence 49, Appl	350	15	1.8	1333	3	US-09-221-235-15	Sequence 15, Appl
C 279	15	1.8	711	4	US-09-299-268-49	Sequence 49, Appl	351	15	1.8	1333	3	US-09-221-235-15	Sequence 15, Appl
C 280	15	1.8	720	3	US-09-479-309-1	Sequence 1, Appl	352	15	1.8	1333	3	US-09-221-235-15	Sequence 15, Appl
C 281	15	1.8	720	3	US-09-479-309-4	Sequence 4, Appl	353	15	1.8	1333	3	US-09-221-235-15	Sequence 15, Appl
C 282	15	1.8	720	3	US-09-479-309-7	Sequence 7, Appl	354	15	1.8	1333	4	US-09-163-115-13	Sequence 13, Appl
C 283	15	1.8	723	1	US-08-338-009B-1	Sequence 1, Appl	355	15	1.8	1333	4	US-09-221-235-15	Sequence 15, Appl
C 284	15	1.8	723	1	US-08-597-583-1	Sequence 1, Appl	356	15	1.8	1333	4	US-09-593-553-13	Sequence 13, Appl
C 285	15	1.8	723	5	PCT-US95-14717-1	Sequence 1, Appl	C 357	15	1.8	1333	4	US-09-221-237-13	Sequence 13, Appl
C 286	15	1.8	732	2	US-08-860-882A-29	Sequence 29, Appl	C 358	15	1.8	1362	6	5290690-8	Patent No. 5290690
C 287	15	1.8	732	4	US-09-437-054A-15	Sequence 15, Appl	C 359	15	1.8	1470	4	US-09-134-001C-2404	Sequence 2404, Ap
C 288	15	1.8	732	4	US-09-011-769A-26	Sequence 26, Appl	C 360	15	1.8	1482	1	US-08-330-121B-1	Sequence 1, Appl
C 289	15	1.8	737	1	US-08-602-262-1	Sequence 1, Appl	C 361	15	1.8	1482	5	PCT-US95-13820-1	Sequence 1, Appl
C 290	15	1.8	737	1	US-08-602-262-2	Sequence 2, Appl	362	15	1.8	1514	1	US-08-434-255-21	Sequence 21, Appl
C 291	15	1.8	737	1	US-08-602-262-3	Sequence 3, Appl	363	15	1.8	1514	1	US-08-459-967-21	Sequence 21, Appl
C 292	15	1.8	737	3	US-09-004-716-1	Sequence 1, Appl	364	15	1.8	1514	1	US-08-460-327-21	Sequence 21, Appl
C 293	15	1.8	759	4	US-09-134-001C-1010	Sequence 1010, Ap	365	15	1.8	1514	1	US-08-459-967-21	Sequence 21, Appl
C 294	15	1.8	759	4	US-09-134-001C-661	Sequence 661, App	C 366	15	1.8	1614	4	US-08-684-932A-35	Sequence 35, Appl
C 295	15	1.8	795	3	US-08-904-284-2	Sequence 2, Appl	C 367	15	1.8	1619	1	US-08-208-007A-1	Sequence 1, Appl
C 296	15	1.8	819	1	US-08-309-182B-2	Sequence 2, Appl	C 368	15	1.8	1619	4	US-08-915-095A-1	Sequence 1, Appl
C 297	15	1.8	850	3	US-09-046-086-5	Sequence 5, Appl	C 369	15	1.8	1619	4	US-08-798-096-1	Sequence 1, Appl
C 298	15	1.8	895	4	US-09-280-116-87	Sequence 87, Appl	C 370	15	1.8	1619	4	US-08-798-095A-1	Sequence 1, Appl
C 299	15	1.8	897	1	US-08-338-009B-3	Sequence 3, Appl	C 371	15	1.8	1619	5	PCT-US94-04781-1	Sequence 1, Appl
C 300	15	1.8	897	1	US-08-597-583-3	Sequence 3, Appl	C 372	15	1.8	1669	3	US-08-964-308-3	Sequence 3, Appl
C 301	15	1.8	897	5	PCT-US95-14717-3	Sequence 3, Appl	C 373	15	1.8	1669	3	US-08-964-313-3	Sequence 3, Appl
C 302	15	1.8	926	4	US-08-936-165A-118	Sequence 118, App	C 374	15	1.8	1669	4	US-09-069-138-3	Sequence 3, Appl
C 303	15	1.8	958	2	US-08-757-046A-5	Sequence 5, Appl	375	15	1.8	1688	2	US-08-649-619B-2	Sequence 2, Appl
C 304	15	1.8	958	3	US-09-447-208-5	Sequence 5, Appl	376	15	1.8	1692	3	US-08-535-057A-1	Sequence 1, Appl
C 305	15	1.8	958	3	US-09-135-988-5	Sequence 5, Appl	377	15	1.8	1723	2	US-08-804-699-1	Sequence 1, Appl
C 306	15	1.8	958	4	US-09-277-716-5	Sequence 5, Appl	378	15	1.8	1740	2	US-07-969-213-1	Sequence 1, Appl
C 307	15	1.8	958	4	US-08-597-274A-5	Sequence 5, Appl	379	15	1.8	1764	2	US-08-642-684-1	Sequence 1, Appl
C 308	15	1.8	958	4	US-08-908-909-5	Sequence 5, Appl	380	15	1.8	1764	4	US-09-157-397-1	Sequence 1, Appl
C 309	15	1.8	958	4	US-08-609-161B-5	Sequence 5, Appl	381	15	1.8	1764	4	US-09-157-397-3	Sequence 3, Appl
C 310	15	1.8	958	4	US-08-990-103-5	Sequence 5, Appl	382	15	1.8	1764	4	US-09-517-871-21	Sequence 21, Appl
C 311	15	1.8	1001	4	US-09-641-638-370	Sequence 370, App	383	15	1.8	1770	2	US-08-642-684-3	Sequence 3, Appl
C 312	15	1.8	1001	4	US-09-641-638-371	Sequence 371, App	C 384	15	1.8	1776	1	US-08-298-426-3	Sequence 3, Appl
C 313	15	1.8	1001	4	US-09-641-638-372	Sequence 372, App	385	15	1.8	1813	5	PCT-US94-12883-3	Sequence 3, Appl
C 314	15	1.8	1017	4	US-08-793-634B-6	Sequence 6, Appl	386	15	1.8	1820	6	5223424-12	Patent No. 5223424
C 315	15	1.8	1064	1	US-08-378-588-15	Sequence 15, Appl	387	15	1.8	1898	1	US-08-342-411A-1	Sequence 1, Appl
C 316	15	1.8	1064	2	US-08-811-094-15	Sequence 15, Appl	388	15	1.8	1923	1	US-08-077-939-14	Sequence 14, Appl
C 317	15	1.8	1064	5	PCT-US94-11121-15	Sequence 15, Appl	389	15	1.8	1923	1	US-08-461-939-14	Sequence 14, Appl
C 318	15	1.8	1068	4	US-09-390-131-2	Sequence 2, Appl	390	15	1.8	1923	1	US-08-461-939-14	Sequence 14, Appl
C 319	15	1.8	1119	4	US-08-961-527-287	Sequence 287, App	391	15	1.8	1931	4	US-08-868-435-28	Sequence 28, Appl

393	1.8	1931	4	US-08-744-231-28	Sequence 28, Appl	466	15	1.8	8959	2	US-08-920-828-1	Sequence 1, Appl
394	1.8	1931	4	US-09-044-718-4	Sequence 4, Appl	467	15	1.8	10607	1	US-08-078-090-3	Sequence 3, Appl
395	1.8	1974	1	US-08-077-939-16	Sequence 16, Appl	468	15	1.8	10906	4	US-08-961-259-63	Sequence 61, Appl
396	1.8	1974	1	US-08-461-599-16	Sequence 16, Appl	C 469	15	1.8	11864	4	US-08-961-527-61	Sequence 61, Appl
397	1.8	1974	1	US-08-461-621-16	Sequence 16, Appl	C 470	15	1.8	13674	2	US-08-852-807-1	Sequence 1, Appl
398	1.8	1974	1	US-08-465-334-16	Sequence 16, Appl	C 471	15	1.8	14176	1	US-08-307-499-1	Sequence 1, Appl
399	1.8	1979	2	US-08-649-6198-1	Sequence 1, Appl	C 472	15	1.8	14176	1	US-08-307-499-1	Sequence 1, Appl
400	1.8	2030	1	US-08-330-518-1	Sequence 1, Appl	C 473	15	1.8	14176	1	US-08-307-499-14	Sequence 14, Appl
401	1.8	2030	1	US-08-330-283-1	Sequence 1, Appl	C 474	15	1.8	14176	1	US-08-307-499-14	Sequence 14, Appl
402	1.8	2030	1	US-08-646-248-1	Sequence 1, Appl	C 475	15	1.8	14176	4	US-09-299-268-1	Sequence 1, Appl
403	1.8	2030	5	PCT-US95-13924-1	Sequence 1, Appl	C 476	15	1.8	14176	4	US-09-299-268-1	Sequence 1, Appl
404	1.8	2030	5	PCT-US95-13931-1	Sequence 1, Appl	C 477	15	1.8	14176	4	US-09-299-268-14	Sequence 14, Appl
405	1.8	2062	1	US-08-073-383-5	Sequence 5, Appl	C 478	15	1.8	14176	4	US-09-299-268-14	Sequence 14, Appl
406	1.8	2062	3	US-08-328-239A-3	Sequence 3, Appl	C 479	15	1.8	18702	4	US-08-961-527-7	Sequence 7, Appl
407	1.8	2062	5	PCT-US94-06365-5	Sequence 5, Appl	C 480	15	1.8	19718	4	US-08-961-527-99	Sequence 99, Appl
408	1.8	2062	5	PCT-US95-13661-3	Sequence 3, Appl	C 481	15	1.8	20986	4	US-08-961-527-54	Sequence 54, Appl
409	1.8	2062	6	5223424-3	Patent No. 5223424	C 482	15	1.8	26700	1	US-08-472-217-1	Sequence 1, Appl
410	1.8	2121	4	US-09-117-250-2	Sequence 2, Appl	C 483	15	1.8	26700	2	US-08-488-199-5	Sequence 5, Appl
C 411	1.8	2178	3	US-09-034-916-1	Sequence 1, Appl	C 484	15	1.8	26700	3	US-08-760-534A-1	Sequence 1, Appl
C 412	1.8	2409	4	US-09-484-970B-101	Sequence 101, Appl	C 485	15	1.8	28720	4	US-09-341-587-7	Sequence 7, Appl
C 413	1.8	2481	2	US-08-630-118A-1	Sequence 1, Appl	C 486	15	1.8	43950	4	US-09-735-934A-3	Sequence 3, Appl
C 414	1.8	2481	2	US-08-838-399-1	Sequence 1, Appl	C 487	15	1.8	49795	4	US-09-453-702B-60	Sequence 60, Appl
C 415	1.8	2481	4	US-09-235-839-1	Sequence 1, Appl	C 488	15	1.8	81001	4	US-09-750-580-1	Sequence 1, Appl
C 416	1.8	2481	4	US-09-327-035-1	Sequence 1, Appl	C 489	15	1.8	81001	4	US-09-750-580-1	Sequence 1, Appl
C 417	1.8	2515	1	US-08-061-465-3	Sequence 3, Appl	C 490	15	1.8	98444	4	US-09-791-211-10	Sequence 10, Appl
C 418	1.8	2604	2	US-08-630-118A-3	Sequence 3, Appl	C 491	15	1.8	99500	4	US-09-798-096-10	Sequence 10, Appl
C 419	1.8	2604	2	US-08-838-399-3	Sequence 3, Appl	C 492	15	1.8	112132	4	US-09-741-150-3	Sequence 3, Appl
C 420	1.8	2604	4	US-09-235-839-3	Sequence 3, Appl	C 493	15	1.8	152331	3	US-09-128-155-16	Sequence 16, Appl
C 421	1.8	2604	4	US-09-327-035-3	Sequence 3, Appl	C 494	14	1.7	17	20	US-08-584-040-2257	Sequence 2257, Ap
C 422	1.8	2631	1	US-08-510-215A-1	Sequence 1, Appl	C 495	14	1.7	20	2	US-08-639-501-104	Sequence 104, App
C 423	1.8	2631	4	US-09-517-871-19	Sequence 19, Appl	C 496	14	1.7	20	3	US-09-044-946-104	Sequence 104, App
C 424	1.8	2661	4	US-09-134-001C-1659	Sequence 1659, Ap	C 497	14	1.7	20	3	US-09-044-908-104	Sequence 104, App
C 425	1.8	2799	3	US-09-461-697-359	Sequence 359, App	C 498	14	1.7	20	4	US-08-657-452A-36	Sequence 36, Appl
C 426	1.8	2799	1	US-08-212-188-1	Sequence 1, Appl	C 499	14	1.7	24	1	US-08-470-179-200	Sequence 200, App
C 427	1.8	2799	5	PCT-US95-02708-1	Sequence 1, Appl	C 500	14	1.7	34	1	US-08-387-315A-2	Sequence 2, Appl
C 428	1.8	2799	5	US-09-517-871-19	Sequence 1, Appl	C 501	14	1.7	34	2	US-08-754-559-2	Sequence 2, Appl
C 429	1.8	2973	4	US-08-451-715A-7	Sequence 7, Appl	C 502	14	1.7	35	3	US-08-961-083-302	Sequence 302, App
C 430	1.8	3025	4	US-09-221-017B-694	Sequence 694, App	C 503	14	1.7	36	3	US-08-966-269-10	Sequence 10, Appl
C 431	1.8	3050	2	US-09-031-442A-21	Sequence 21, Appl	C 504	14	1.7	36	4	US-09-436-183A-10	Sequence 10, Appl
C 432	1.8	3050	4	US-09-258-377-21	Sequence 21, Appl	C 505	14	1.7	46	4	US-09-171-878-12	Sequence 12, Appl
C 433	1.8	3168	1	US-07-723-002C-7	Sequence 7, Appl	C 506	14	1.7	71	1	US-08-294-189-21	Sequence 21, Appl
C 434	1.8	317	2	US-08-633-476-1	Sequence 1, Appl	C 507	14	1.7	81	1	US-08-253-879C-60	Sequence 60, Appl
C 435	1.8	3630	4	US-09-221-017B-943	Sequence 943, App	C 508	14	1.7	81	2	US-08-452-164A-60	Sequence 60, Appl
C 436	1.8	3707	1	US-08-118-103A-1	Sequence 1, Appl	C 509	14	1.7	81	4	US-08-976-183A-39	Sequence 39, Appl
C 437	1.8	3747	4	US-09-213-293D-2	Sequence 2, Appl	C 510	14	1.7	97	4	US-08-020-956-84	Sequence 64, Appl
C 438	1.8	4078	2	US-08-960-022-3	Sequence 3, Appl	C 511	14	1.7	97	4	US-09-030-607-64	Sequence 64, Appl
C 439	1.8	4181	4	US-09-393-569-1	Sequence 1, Appl	C 512	14	1.7	97	4	US-09-605-785-64	Sequence 64, Appl
C 440	1.8	4285	4	US-09-410-464-1	Sequence 1, Appl	C 513	14	1.7	97	4	US-09-439-313-64	Sequence 64, Appl
C 441	1.8	4886	4	US-09-533-220A-4	Sequence 4, Appl	C 514	14	1.7	97	4	US-09-352-616A-64	Sequence 64, Appl
C 442	1.8	5134	2	US-08-635-121-1	Sequence 1, Appl	C 515	14	1.7	97	4	US-09-232-149A-64	Sequence 64, Appl
C 443	1.8	5261	1	US-08-045-806-3	Sequence 3, Appl	C 516	14	1.7	110	1	US-08-248-474-10	Sequence 10, Appl
C 444	1.8	5261	1	US-08-366-051B-3	Sequence 3, Appl	C 517	14	1.7	110	3	US-08-756-849-10	Sequence 10, Appl
C 445	1.8	5295	4	US-09-221-017B-27	Sequence 27, Appl	C 518	14	1.7	169	4	US-08-651-155B-203	Sequence 203, App
C 446	1.8	5595	1	US-07-841-655-1	Sequence 1, Appl	C 519	14	1.7	210	5	PCT-US96-05320A-416	Sequence 416, App
C 447	1.8	5595	1	US-07-635-965C-1	Sequence 1, Appl	C 520	14	1.7	228	1	US-08-307-499-42	Sequence 42, Appl
C 448	1.8	6091	4	US-09-221-017B-803	Sequence 803, App	C 521	14	1.7	228	3	US-08-688-988-25	Sequence 25, Appl
C 449	1.8	6933	4	US-08-361-521-195	Sequence 195, App	C 522	14	1.7	228	4	US-08-299-268-42	Sequence 42, Appl
C 450	1.8	6933	1	US-08-061-465-1	Sequence 1, Appl	C 523	14	1.7	231	4	US-09-134-001C-2237	Sequence 2237, Ap
C 451	1.8	7493	1	US-08-212-133A-7	Sequence 7, Appl	C 524	14	1.7	237	4	US-09-134-001C-810	Sequence 810, App
C 452	1.8	7493	2	US-08-474-503-5	Sequence 5, Appl	C 525	14	1.7	239	1	US-08-253-155A-50	Sequence 50, Appl
C 453	1.8	7493	2	US-08-670-707A-5	Sequence 5, Appl	C 526	14	1.7	253	4	US-08-651-155B-204	Sequence 204, App
C 454	1.8	7493	4	US-09-037-601-5	Sequence 5, Appl	C 527	14	1.7	280	1	US-08-171-385-23	Sequence 23, Appl
C 455	1.8	7493	5	PCT-US94-13200-5	Sequence 5, Appl	C 528	14	1.7	280	3	US-08-361-441B-23	Sequence 23, Appl
C 456	1.8	8543	3	US-08-496-944-1	Sequence 1, Appl	C 529	14	1.7	282	4	US-08-651-155B-139	Sequence 139, App
C 457	1.8	8855	2	US-08-542-003-1	Sequence 1, Appl	C 530	14	1.7	301	4	US-09-605-785-272	Sequence 272, App
C 458	1.8	8855	2	US-08-322-760A-1	Sequence 1, Appl	C 531	14	1.7	301	4	US-09-439-313-272	Sequence 272, App
C 459	1.8	8855	4	US-09-236-949-1	Sequence 1, Appl	C 532	14	1.7	301	4	US-09-352-616A-272	Sequence 272, App
C 460	1.8	8898	4	US-08-361-521-69	Sequence 69, Appl	C 533	14	1.7	307	4	US-09-232-149A-272	Sequence 272, App
C 461	1.8	8959	1	US-08-920-827-1	Sequence 1, Appl	C 534	14	1.7	310	4	US-08-171-385-8	Sequence 8, Appl
C 462	1.8	8959	1	US-08-920-827-1	Sequence 1, Appl	C 535	14	1.7	310	3	US-08-361-441B-8	Sequence 8, Appl
C 463	1.8	8959	1	US-08-921-177-1	Sequence 1, Appl	C 536	14	1.7	313	1	US-08-171-385-15	Sequence 15, Appl
C 464	1.8	8959	1	US-08-362-577C-1	Sequence 1, Appl	C 537	14	1.7	313	3	US-08-361-441B-15	Sequence 15, Appl
C 465	1.8	8959	1			C 538	14	1.7	313	3		

C 539	14	1.7	315	4	US-09-134-001C-2743	Sequence 2743, Ap	612	14	1.7	864	3	US-08-733-230-1	Sequence 1, Appl
C 540	14	1.7	329	1	US-08-171-385-21	Sequence 21, Appl	613	14	1.7	864	4	US-08-953-326-1	Sequence 12, Appl
C 541	14	1.7	329	3	US-08-361-441B-21	Sequence 21, Appl	614	14	1.7	864	4	US-08-953-326-12	Sequence 12, Appl
C 542	14	1.7	339	4	US-09-180-077-3	Sequence 3, Appl	615	14	1.7	872	4	US-08-998-417-487	Sequence 487, App
C 543	14	1.7	339	4	US-09-180-077-8	Sequence 8, Appl	616	14	1.7	892	4	US-09-323-427-16	Sequence 16, Appl
C 544	14	1.7	358	3	US-08-961-083-103	Sequence 103, App	617	14	1.7	892	4	US-09-323-427-18	Sequence 18, Appl
C 545	14	1.7	360	4	US-08-961-083-43	Sequence 43, Appl	618	14	1.7	892	4	US-09-812-642-16	Sequence 16, Appl
C 546	14	1.7	360	4	US-08-352-902D-43	Sequence 43, Appl	619	14	1.7	892	4	US-09-812-642-18	Sequence 18, Appl
C 547	14	1.7	384	4	US-08-943-731-49	Sequence 49, Appl	620	14	1.7	892	4	US-09-668-680-5	Sequence 5, Appl
C 548	14	1.7	384	4	US-09-134-001C-59	Sequence 59, Appl	621	14	1.7	915	4	US-09-134-001C-2316	Sequence 2316, Ap
C 549	14	1.7	385	3	US-08-733-708C-42	Sequence 42, Appl	622	14	1.7	915	4	US-08-483-639-2	Sequence 2, Appl
C 550	14	1.7	393	4	US-09-134-001C-354	Sequence 354, App	623	14	1.7	925	4	US-09-180-077-1	Sequence 1, Appl
C 551	14	1.7	395	4	US-08-936-165A-61	Sequence 61, Appl	624	14	1.7	925	4	US-08-943-731-51	Sequence 51, Appl
C 552	14	1.7	396	1	US-08-470-179-61	Sequence 61, Appl	625	14	1.7	940	1	US-07-691-079C-1	Sequence 1, Appl
C 553	14	1.7	396	4	US-08-887-534A-21	Sequence 21, Appl	626	14	1.7	940	1	US-08-433-037-1	Sequence 1, Appl
C 554	14	1.7	410	4	US-09-534-638-8	Sequence 8, Appl	627	14	1.7	967	4	US-08-960-780-47	Sequence 47, Appl
C 555	14	1.7	423	2	US-08-619-708A-1	Sequence 1, Appl	628	14	1.7	967	4	US-09-073-898-47	Sequence 47, Appl
C 556	14	1.7	450	4	US-09-397-787-173	Sequence 173, App	629	14	1.7	971	1	US-08-220-606B-30	Sequence 30, Appl
C 557	14	1.7	468	4	US-09-134-001C-2729	Sequence 2729, Ap	630	14	1.7	972	1	US-07-915-934-1	Sequence 1, Appl
C 558	14	1.7	475	4	US-09-221-017B-422	Sequence 422, App	631	14	1.7	972	1	US-08-325-743-1	Sequence 1, Appl
C 559	14	1.7	498	4	US-09-328-111-199	Sequence 199, App	632	14	1.7	972	4	US-08-960-780-48	Sequence 48, Appl
C 560	14	1.7	498	4	US-09-347-801-7	Sequence 7, Appl	633	14	1.7	972	4	US-09-073-898-48	Sequence 48, Appl
C 561	14	1.7	500	4	US-09-330-611-26	Sequence 26, Appl	634	14	1.7	973	4	US-09-180-077-2	Sequence 2, Appl
C 562	14	1.7	501	4	US-09-134-001C-2536	Sequence 2536, Ap	635	14	1.7	987	4	US-08-961-527-215	Sequence 215, App
C 563	14	1.7	508	4	US-09-221-017B-535	Sequence 535, App	636	14	1.7	995	4	US-09-556-877-115	Sequence 115, App
C 564	14	1.7	521	4	US-09-404-879A-34	Sequence 34, App	637	14	1.7	995	4	US-09-620-412C-115	Sequence 115, App
C 565	14	1.7	523	4	US-09-221-017B-1076	Sequence 1076, Ap	638	14	1.7	1001	4	US-09-641-638-135	Sequence 135, App
C 566	14	1.7	537	1	US-08-176-620A-9	Sequence 9, Appl	639	14	1.7	1001	4	US-09-641-638-136	Sequence 136, App
C 567	14	1.7	537	2	US-08-461-985-9	Sequence 9, Appl	640	14	1.7	1001	4	US-09-641-638-137	Sequence 137, App
C 568	14	1.7	537	4	US-08-932-787B-17	Sequence 17, Appl	641	14	1.7	1001	4	US-09-641-638-138	Sequence 138, App
C 569	14	1.7	537	4	US-08-932-012C-17	Sequence 17, Appl	642	14	1.7	1001	4	US-09-641-638-354	Sequence 354, App
C 570	14	1.7	537	4	US-08-888-818C-17	Sequence 17, Appl	643	14	1.7	1001	4	US-09-641-638-396	Sequence 396, App
C 571	14	1.7	541	4	US-09-404-879A-11	Sequence 11, Appl	644	14	1.7	1010	4	US-09-011-526-3	Sequence 3, Appl
C 572	14	1.7	563	4	US-08-961-810-24	Sequence 24, Appl	645	14	1.7	1018	3	US-09-029-267-18	Sequence 18, Appl
C 573	14	1.7	563	4	US-08-352-902D-24	Sequence 24, Appl	646	14	1.7	1050	1	US-08-180-209B-16	Sequence 16, Appl
C 574	14	1.7	570	4	US-09-328-111-461	Sequence 461, App	647	14	1.7	1050	1	US-08-385-745-16	Sequence 16, Appl
C 575	14	1.7	575	2	US-08-966-316-9	Sequence 9, Appl	648	14	1.7	1050	4	US-08-485-388-16	Sequence 16, Appl
C 576	14	1.7	607	4	US-09-385-982-372	Sequence 372, App	649	14	1.7	1050	4	US-08-474-853-16	Sequence 16, Appl
C 577	14	1.7	613	4	US-09-149-476-15	Sequence 15, Appl	650	14	1.7	1050	4	US-09-166-205B-16	Sequence 16, Appl
C 578	14	1.7	639	4	US-09-221-017B-239	Sequence 239, App	651	14	1.7	1050	5	PCT-US94-02629-16	Sequence 16, Appl
C 579	14	1.7	660	4	US-09-417-822-41	Sequence 41, Appl	652	14	1.7	1057	1	US-08-470-202-44	Sequence 44, Appl
C 580	14	1.7	661	2	US-08-528-878B-37	Sequence 37, Appl	653	14	1.7	1057	2	US-08-471-770-44	Sequence 44, Appl
C 581	14	1.7	662	4	US-08-936-165A-86	Sequence 86, Appl	654	14	1.7	1057	4	US-08-468-059-44	Sequence 44, Appl
C 582	14	1.7	663	4	US-08-998-416-933	Sequence 933, App	655	14	1.7	1057	4	US-09-109-916-44	Sequence 44, Appl
C 583	14	1.7	677	1	US-08-211-942-8	Sequence 8, Appl	656	14	1.7	1057	4	US-09-109-916-45	Sequence 45, Appl
C 584	14	1.7	677	4	US-08-998-416-1092	Sequence 1092, Ap	657	14	1.7	1061	4	US-09-247-155-57	Sequence 57, Appl
C 585	14	1.7	689	4	US-09-152-060-13	Sequence 13, Appl	658	14	1.7	1069	4	US-09-071-035-7	Sequence 7, Appl
C 586	14	1.7	689	5	PCT-US95-06406A-9	Sequence 9, Appl	659	14	1.7	1083	4	US-09-134-001C-1520	Sequence 1520, Ap
C 587	14	1.7	699	4	US-09-247-155-40	Sequence 40, Appl	660	14	1.7	1097	1	US-08-021-608B-5	Sequence 5, Appl
C 588	14	1.7	699	5	PCT-US95-13335-2	Sequence 2, Appl	661	14	1.7	1097	1	US-08-726-160-5	Sequence 5, Appl
C 589	14	1.7	719	4	US-08-998-416-731	Sequence 731, App	662	14	1.7	1097	5	PCT-US94-01783-5	Sequence 5, Appl
C 590	14	1.7	720	3	US-09-479-309-5	Sequence 5, Appl	663	14	1.7	1105	3	US-08-961-083-75	Sequence 75, Appl
C 591	14	1.7	720	4	US-09-118-442-30	Sequence 30, Appl	664	14	1.7	1119	3	US-08-911-321-2	Sequence 2, Appl
C 592	14	1.7	720	4	US-09-677-064-30	Sequence 30, Appl	665	14	1.7	1123	1	US-08-105-483-290	Sequence 290, App
C 593	14	1.7	722	4	US-09-180-077-5	Sequence 5, Appl	666	14	1.7	1125	1	US-08-709-209-290	Sequence 290, App
C 594	14	1.7	731	1	US-08-451-405A-2	Sequence 2, Appl	667	14	1.7	1125	1	US-08-458-101-290	Sequence 290, App
C 595	14	1.7	732	1	US-08-211-942-6	Sequence 6, Appl	668	14	1.7	1128	1	US-08-465-590-7	Sequence 7, Appl
C 596	14	1.7	770	4	US-09-180-077-10	Sequence 10, Appl	669	14	1.7	1128	4	US-08-283-300A-6	Sequence 6, Appl
C 597	14	1.7	783	4	US-09-134-001C-2035	Sequence 2035, Ap	670	14	1.7	1128	4	US-08-711-417C-7	Sequence 7, Appl
C 598	14	1.7	810	4	US-09-134-001C-2360	Sequence 2360, Ap	671	14	1.7	1128	5	PCT-US95-09348-6	Sequence 6, Appl
C 599	14	1.7	813	4	US-09-693-147-3	Sequence 3, Appl	672	14	1.7	1133	4	US-09-395-674B-5	Sequence 5, Appl
C 600	14	1.7	817	4	US-09-318-448-7	Sequence 7, Appl	673	14	1.7	1137	4	US-09-553-498-9	Sequence 9, Appl
C 601	14	1.7	819	4	US-09-605-785-571	Sequence 571, App	674	14	1.7	1137	4	US-09-618-869-9	Sequence 9, Appl
C 602	14	1.7	821	4	US-08-998-416-541	Sequence 541, App	675	14	1.7	1142	1	US-08-006-082A-5	Sequence 5, Appl
C 603	14	1.7	822	4	US-09-134-001C-804	Sequence 804, App	676	14	1.7	1146	4	US-09-134-001C-906	Sequence 906, App
C 604	14	1.7	825	4	US-09-280-839-1	Sequence 1, Appl	677	14	1.7	1158	1	US-08-891-254-4	Sequence 4, Appl
C 605	14	1.7	825	4	US-09-134-001C-1452	Sequence 1452, Ap	678	14	1.7	1158	2	US-08-819-539-4	Sequence 4, Appl
C 606	14	1.7	834	4	US-09-134-001C-994	Sequence 994, App	679	14	1.7	1158	5	PCT-US93-06243-4	Sequence 4, Appl
C 607	14	1.7	836	4	US-09-247-373B-49	Sequence 49, Appl	680	14	1.7	1158	5	PCT-US96-08819-4	Sequence 4, Appl
C 608	14	1.7	838	4	US-09-122-400B-21	Sequence 21, Appl	681	14	1.7	1160	1	US-08-006-082A-4	Sequence 4, Appl
C 609	14	1.7	845	4	US-09-292-858B-9	Sequence 9, Appl	682	14	1.7	1161	4	US-09-134-001C-972	Sequence 972, App
C 610	14	1.7	853	3	US-09-081-180-1	Sequence 1, Appl	683	14	1.7	1161	4	US-09-071-035-5	Sequence 5, Appl
C 611	14	1.7	853	3	US-09-040-786-1	Sequence 1, Appl	684	14	1.7	1164	4	US-09-134-001C-1245	Sequence 1245, Ap

685	14	1.7	1172	4	US-09-140-084-21	Sequence 21, Appl	c 758	14	1.7	1465	4	US-09-004-838-100	Sequence 100, App
686	14	1.7	1172	4	US-09-724-297-20	Sequence 20, Appl	c 759	14	1.7	1466	4	US-09-004-838-33	Sequence 33, Appl
687	14	1.7	1203	4	US-09-434-727-9	Sequence 9, Appl	760	14	1.7	1478	1	US-08-090-523-27	Sequence 27, Appl
c 688	14	1.7	1209	1	US-08-726-227-2	Sequence 2, Appl	761	14	1.7	1478	1	US-08-398-627-37	Sequence 37, Appl
c 689	14	1.7	1209	2	US-08-200-724A-4	Sequence 4, Appl	762	14	1.7	1478	1	US-08-406-857-1	Sequence 1, Appl
c 690	14	1.7	1209	4	US-08-851-376A-4	Sequence 4, Appl	763	14	1.7	1478	1	US-08-596-024-3	Sequence 3, Appl
691	14	1.7	1224	2	US-08-210-762E-19	Sequence 19, Appl	764	14	1.7	1478	4	US-09-020-818-3	Sequence 3, Appl
692	14	1.7	1224	4	US-09-106-075A-18	Sequence 18, Appl	765	14	1.7	1478	4	US-08-907-740-3	Sequence 3, Appl
693	14	1.7	1236	4	US-08-964-652-1	Sequence 1, Appl	766	14	1.7	1478	5	PCT-US94-07072-1	Sequence 1, Appl
694	14	1.7	1241	1	US-07-593-657-6	Sequence 6, Appl	c 767	14	1.7	1490	3	US-08-605-150A-5	Sequence 5, Appl
695	14	1.7	1241	4	US-08-942-012B-3	Sequence 3, Appl	c 768	14	1.7	1500	4	US-08-875-847B-3	Sequence 3, Appl
696	14	1.7	1251	4	US-09-330-611-1	Sequence 1, Appl	c 769	14	1.7	1500	4	US-09-378-842-3	Sequence 3, Appl
697	14	1.7	1251	4	US-09-330-611-7	Sequence 7, Appl	c 770	14	1.7	1509	4	US-09-134-001C-1674	Sequence 1674, Ap
c 698	14	1.7	1260	1	US-08-467-933-7	Sequence 7, Appl	771	14	1.7	1512	2	US-08-270-581-1	Sequence 1, Appl
699	14	1.7	1261	4	US-08-858-207A-151	Sequence 151, App	c 772	14	1.7	1512	3	US-08-945-994-4	Sequence 4, Appl
700	14	1.7	1264	1	US-08-006-082A-1	Sequence 1, Appl	773	14	1.7	1512	4	US-09-146-893-1	Sequence 1, Appl
701	14	1.7	1276	4	US-08-858-207A-109	Sequence 109, App	774	14	1.7	1528	4	US-09-345-214-6	Sequence 6, Appl
c 702	14	1.7	1284	5	PCT-US96-03940-28	Sequence 28, Appl	c 775	14	1.7	1551	4	US-08-711-417C-165	Sequence 165, App
c 703	14	1.7	1287	5	PCT-US93-06243-3	Sequence 3, Appl	c 776	14	1.7	1568	1	US-08-589-446-7	Sequence 7, Appl
c 704	14	1.7	1288	2	US-08-200-724A-3	Sequence 3, Appl	c 777	14	1.7	1568	1	US-08-444-882-7	Sequence 7, Appl
c 705	14	1.7	1288	2	US-09-030-270A-4	Sequence 4, Appl	c 778	14	1.7	1568	3	US-08-389-459A-7	Sequence 7, Appl
c 706	14	1.7	1288	4	US-08-851-376A-3	Sequence 3, Appl	c 779	14	1.7	1568	3	US-08-987-867A-7	Sequence 7, Appl
c 707	14	1.7	1288	4	US-08-984-207-4	Sequence 4, Appl	780	14	1.7	1575	1	US-08-090-523-7	Sequence 7, Appl
c 708	14	1.7	1288	4	US-09-013-587-4	Sequence 4, Appl	781	14	1.7	1575	1	US-08-398-627-7	Sequence 7, Appl
c 709	14	1.7	1289	1	US-08-644-664B-13	Sequence 13, Appl	782	14	1.7	1575	1	US-08-406-858-7	Sequence 7, Appl
c 710	14	1.7	1289	2	US-08-761-277A-13	Sequence 13, Appl	783	14	1.7	1575	5	PCT-US91-04036-7	Sequence 7, Appl
c 711	14	1.7	1296	1	US-08-465-590-4	Sequence 4, Appl	784	14	1.7	1575	5	PCT-US94-05275-7	Sequence 7, Appl
c 712	14	1.7	1296	4	US-08-283-300A-3	Sequence 3, Appl	c 785	14	1.7	1586	1	US-08-461-244-1	Sequence 1, Appl
c 713	14	1.7	1296	4	US-08-711-417C-4	Sequence 4, Appl	786	14	1.7	1593	1	US-08-307-499-50	Sequence 50, Appl
c 714	14	1.7	1296	5	PCT-US95-09345-3	Sequence 3, Appl	787	14	1.7	1593	4	US-09-299-268-50	Sequence 50, Appl
715	14	1.7	1333	4	US-08-543-246B-15	Sequence 15, Appl	c 788	14	1.7	1596	3	US-08-388-553-651	Sequence 651, App
c 716	14	1.7	1350	4	US-09-134-001C-1572	Sequence 1572, Ap	c 789	14	1.7	1596	3	US-08-488-551B-651	Sequence 651, App
c 717	14	1.7	1354	3	US-08-605-150A-1	Sequence 1, Appl	790	14	1.7	1600	3	US-09-032-372-7	Sequence 7, Appl
718	14	1.7	1356	1	US-07-603-133B-1	Sequence 1, Appl	c 791	14	1.7	1604	4	US-09-260-843-1	Sequence 1, Appl
719	14	1.7	1356	1	US-08-426-169-3	Sequence 3, Appl	c 792	14	1.7	1604	4	US-09-923-654-1	Sequence 1, Appl
720	14	1.7	1356	5	US-09-233-813-3	Sequence 3, Appl	c 793	14	1.7	1611	5	PCT-US93-08743-3	Sequence 3, Appl
721	14	1.7	1356	5	PCT-US95-09470-3	Sequence 3, Appl	c 794	14	1.7	1620	3	US-08-941-445A-20	Sequence 20, Appl
722	14	1.7	1357	3	US-08-089-397A-17	Sequence 17, Appl	795	14	1.7	1667	1	US-08-186-016A-2	Sequence 2, Appl
723	14	1.7	1365	4	US-09-134-001C-710	Sequence 710, App	c 796	14	1.7	1672	4	US-09-495-066-1	Sequence 1, Appl
c 724	14	1.7	1365	3	US-08-945-994-1	Sequence 1, Appl	c 797	14	1.7	1672	4	US-09-134-001C-1341	Sequence 1341, Ap
c 725	14	1.7	1380	1	US-08-465-590-3	Sequence 3, Appl	c 798	14	1.7	1690	1	US-08-276-452A-24	Sequence 24, Appl
c 726	14	1.7	1386	4	US-08-283-300A-2	Sequence 2, Appl	c 799	14	1.7	1690	2	US-08-798-744-24	Sequence 24, Appl
c 727	14	1.7	1386	4	US-08-711-417C-3	Sequence 3, Appl	800	14	1.7	1695	4	US-09-142-623-12	Sequence 12, Appl
c 728	14	1.7	1386	4	US-08-711-417C-3	Sequence 3, Appl	c 801	14	1.7	1696	1	US-08-340-539A-11	Sequence 11, Appl
c 729	14	1.7	1386	5	PCT-US95-09345-2	Sequence 2, Appl	c 802	14	1.7	1696	2	US-08-461-552B-11	Sequence 11, Appl
c 730	14	1.7	1386	5	US-09-484-970B-123	Sequence 123, App	803	14	1.7	1704	3	US-09-125-287-4	Sequence 4, Appl
731	14	1.7	1387	3	US-08-868-594-1	Sequence 1, Appl	c 804	14	1.7	1725	1	US-08-257-073-14	Sequence 14, Appl
732	14	1.7	1387	4	US-08-543-246B-1	Sequence 1, Appl	c 805	14	1.7	1734	4	US-09-484-970B-63	Sequence 63, Appl
733	14	1.7	1387	4	US-09-434-323-1	Sequence 1, Appl	c 806	14	1.7	1751	4	US-09-134-001C-2774	Sequence 2774, Ap
c 734	14	1.7	1390	3	US-08-605-150A-7	Sequence 7, Appl	c 807	14	1.7	1751	4	US-09-004-393B-3	Sequence 3, Appl
c 735	14	1.7	1393	3	US-08-826-246-5	Sequence 5, Appl	c 808	14	1.7	1752	3	US-08-941-445A-12	Sequence 12, Appl
c 736	14	1.7	1393	3	US-08-944-495-5	Sequence 5, Appl	c 809	14	1.7	1791	5	PCT-US96-03940-6	Sequence 6, Appl
c 737	14	1.7	1393	3	US-09-126-640-3	Sequence 3, Appl	810	14	1.7	1803	1	US-08-021-608D-7	Sequence 7, Appl
c 738	14	1.7	1393	4	US-08-925-588-5	Sequence 5, Appl	811	14	1.7	1803	1	US-08-726-160-7	Sequence 7, Appl
c 739	14	1.7	1393	4	US-09-288-292A-3	Sequence 3, Appl	812	14	1.7	1803	5	PCT-US94-01782-7	Sequence 7, Appl
c 740	14	1.7	1395	3	US-08-834-776A-1	Sequence 1, Appl	813	14	1.7	1810	1	US-07-755-573C-7	Sequence 7, Appl
c 741	14	1.7	1402	4	US-09-196-520-5	Sequence 5, Appl	814	14	1.7	1818	4	US-08-041-236-3	Sequence 3, Appl
c 742	14	1.7	1407	5	PCT-US96-03940-3	Sequence 3, Appl	815	14	1.7	1818	4	US-09-357-206A-6	Sequence 6, Appl
c 743	14	1.7	1410	2	US-08-847-900-2	Sequence 2, Appl	c 816	14	1.7	1822	4	US-09-393-245-1	Sequence 1, Appl
c 744	14	1.7	1415	4	US-09-345-214-11	Sequence 11, Appl	c 817	14	1.7	1822	4	US-09-393-245-3	Sequence 3, Appl
c 745	14	1.7	1419	4	US-09-004-838-102	Sequence 102, App	818	14	1.7	1827	4	US-09-310-275-1	Sequence 1, Appl
c 746	14	1.7	1420	4	US-09-004-838-34	Sequence 34, Appl	819	14	1.7	1831	4	US-09-336-536-15	Sequence 15, Appl
747	14	1.7	1423	1	US-08-596-111B-1	Sequence 1, Appl	820	14	1.7	1837	2	US-08-909-965C-10	Sequence 10, Appl
748	14	1.7	1428	2	US-08-624-601-7	Sequence 7, Appl	821	14	1.7	1848	1	US-08-333-802-1	Sequence 1, Appl
c 749	14	1.7	1436	4	US-09-004-838-104	Sequence 104, App	c 822	14	1.7	1851	4	US-09-134-001C-1781	Sequence 1781, Ap
c 750	14	1.7	1442	4	US-08-936-165A-67	Sequence 67, Appl	823	14	1.7	1862	4	US-09-484-970B-74	Sequence 74, Appl
c 751	14	1.7	1442	4	US-09-004-838-35	Sequence 35, Appl	824	14	1.7	1863	2	US-08-455-073A-3	Sequence 3, Appl
c 752	14	1.7	1451	1	US-07-924-028A-1	Sequence 1, Appl	c 825	14	1.7	1886	6	5256558-1	Patent No. 5256558
c 753	14	1.7	1451	4	US-09-276-531-28	Sequence 28, Appl	826	14	1.7	1888	3	US-09-035-648-20	Sequence 20, Appl
c 754	14	1.7	1454	4	US-09-004-838-130	Sequence 130, App	827	14	1.7	1888	4	US-09-001-951-20	Sequence 20, Appl
c 755	14	1.7	1461	4	US-08-858-207A-44	Sequence 44, Appl	828	14	1.7	1888	4	US-08-818-829-20	Sequence 20, Appl
c 756	14	1.7	1461	5	PCT-US96-03940-27	Sequence 27, Appl	c 829	14	1.7	1897	1	US-08-453-477-1	Sequence 1, Appl
c 757	14	1.7	1464	3	US-08-605-150A-3	Sequence 3, Appl	c 830	14	1.7	1897	1	US-08-453-461-1	Sequence 1, Appl

831	14	1.7	1927	4	US-09-336-536-66	Sequence 66, Appl	c 904	14	1.7	2515	4	US-09-266-417-37	Sequence 37, Appl
832	14	1.7	1939	1	US-07-715-751B-2	Sequence 2, Appl	c 905	14	1.7	2517	4	US-09-534-407-5	Sequence 5, Appl
833	14	1.7	1986	4	US-09-221-017B-150	Sequence 150, App	c 906	14	1.7	2525	4	US-08-294-312B-1	Sequence 1, Appl
834	14	1.7	1990	4	US-08-961-527-232	Sequence 232, App	c 907	14	1.7	2525	4	US-08-468-024B-1	Sequence 1, Appl
835	14	1.7	2008	4	US-09-345-214-12	Sequence 12, Appl	c 908	14	1.7	2530	3	US-08-792-832A-1	Sequence 1, Appl
836	14	1.7	2026	2	US-08-559-492-1	Sequence 1, Appl	c 909	14	1.7	2531	3	US-07-956-483-18	Sequence 18, Appl
837	14	1.7	2034	3	US-08-834-314-1	Sequence 1, Appl	c 910	14	1.7	2534	3	US-09-125-287-3	Sequence 3, Appl
838	14	1.7	2037	4	US-08-913-942-14	Sequence 14, Appl	c 911	14	1.7	2546	3	US-07-916-098A-1	Sequence 1, Appl
839	14	1.7	2042	4	US-08-894-997-39	Sequence 39, Appl	c 912	14	1.7	2571	2	US-09-289-354-1	Sequence 1, Appl
840	14	1.7	2043	2	US-08-398-590A-39	Sequence 39, Appl	c 913	14	1.7	2596	4	US-09-556-877-181	Sequence 181, App
841	14	1.7	2049	1	US-08-465-590-5	Sequence 5, Appl	c 914	14	1.7	2601	4	US-09-620-412C-181	Sequence 181, App
842	14	1.7	2049	4	US-08-283-300A-4	Sequence 4, Appl	c 915	14	1.7	2605	4	US-09-651-656-104	Sequence 104, App
843	14	1.7	2049	4	US-08-711-417C-5	Sequence 5, Appl	c 916	14	1.7	2605	4	US-09-650-855-104	Sequence 104, App
844	14	1.7	2049	5	PCT-US95-09345-4	Sequence 4, Appl	c 917	14	1.7	2622	1	US-08-766-014-23	Sequence 23, Appl
845	14	1.7	2052	4	US-09-620-412C-356	Sequence 356, App	c 918	14	1.7	2623	3	US-09-234-332-6	Sequence 6, Appl
846	14	1.7	2056	4	US-08-158-735A-1	Sequence 1, Appl	c 919	14	1.7	2631	1	US-08-785-071A-1	Sequence 1, Appl
847	14	1.7	2070	4	US-09-382-256-13	Sequence 13, Appl	c 920	14	1.7	2631	3	US-09-012-872-1	Sequence 1, Appl
848	14	1.7	2070	4	US-09-395-115-13	Sequence 13, Appl	c 921	14	1.7	2643	4	US-09-556-877-169	Sequence 169, App
849	14	1.7	2070	4	US-08-436-265-13	Sequence 13, Appl	c 922	14	1.7	2643	4	US-09-620-412C-169	Sequence 169, App
850	14	1.7	2070	4	US-09-679-187-13	Sequence 13, Appl	c 923	14	1.7	2644	4	US-08-472-240A-9	Sequence 9, Appl
851	14	1.7	2079	4	US-09-268-347-25	Sequence 25, Appl	c 924	14	1.7	2652	4	US-09-115-704-1	Sequence 1, Appl
852	14	1.7	2085	4	US-09-556-877-81	Sequence 81, Appl	c 925	14	1.7	2692	1	US-08-036-210-14	Sequence 14, Appl
853	14	1.7	2085	4	US-09-620-412C-81	Sequence 81, Appl	c 926	14	1.7	2692	2	US-08-419-609-14	Sequence 14, Appl
854	14	1.7	2102	4	US-09-318-448-19	Sequence 19, Appl	c 927	14	1.7	2694	1	US-08-147-890-1	Sequence 1, Appl
855	14	1.7	2114	4	US-09-370-838-158	Sequence 158, App	c 928	14	1.7	2712	1	US-08-346-455B-37	Sequence 37, Appl
856	14	1.7	2115	4	US-09-117-860-23	Sequence 23, Appl	c 929	14	1.7	2712	3	US-08-977-221-37	Sequence 37, Appl
857	14	1.7	2126	4	US-08-235-836C-75	Sequence 75, Appl	c 930	14	1.7	2712	4	US-09-483-831B-37	Sequence 37, Appl
858	14	1.7	2127	1	US-08-158-232-50	Sequence 50, Appl	c 931	14	1.7	2712	5	PCT-US95-06613-37	Sequence 37, Appl
859	14	1.7	2172	1	US-08-611-928-50	Sequence 50, Appl	c 932	14	1.7	2713	1	US-08-285-439-1	Sequence 1, Appl
860	14	1.7	2172	3	US-08-173-891-50	Sequence 50, Appl	c 933	14	1.7	2713	4	US-09-074-658-69	Sequence 69, Appl
861	14	1.7	2173	2	US-08-036-210-6	Sequence 6, Appl	c 934	14	1.7	2730	3	US-08-728-122-1	Sequence 1, Appl
862	14	1.7	2173	2	US-08-449-609-6	Sequence 6, Appl	c 935	14	1.7	2751	1	US-08-153-848-23	Sequence 23, Appl
863	14	1.7	2181	4	US-09-299-041-1	Sequence 1, Appl	c 936	14	1.7	2751	3	US-09-299-843A-23	Sequence 23, Appl
864	14	1.7	2184	4	US-09-337-387-4	Sequence 4, Appl	c 937	14	1.7	2751	4	US-09-088-337B-23	Sequence 23, Appl
865	14	1.7	2190	4	US-09-729-995-1	Sequence 1, Appl	c 938	14	1.7	2751	5	PCT-US93-11153-23	Sequence 23, Appl
866	14	1.7	2211	4	US-09-134-001C-818	Sequence 818, App	c 939	14	1.7	2788	4	US-09-347-878-45	Sequence 45, Appl
867	14	1.7	2239	4	US-09-196-390-1	Sequence 1, Appl	c 940	14	1.7	2797	2	US-09-021-323-2	Sequence 2, Appl
868	14	1.7	2246	4	US-09-117-860-17	Sequence 17, Appl	c 941	14	1.7	2803	3	US-08-949-588-1	Sequence 1, Appl
869	14	1.7	2250	1	US-08-483-639-1	Sequence 1, Appl	c 942	14	1.7	2815	4	US-09-276-531-127	Sequence 127, App
870	14	1.7	2293	3	US-09-125-287-5	Sequence 5, Appl	c 943	14	1.7	2817	5	PCT-US93-05944-1	Sequence 1, Appl
871	14	1.7	2309	1	US-08-036-210-10	Sequence 10, Appl	c 944	14	1.7	2823	4	US-09-513-783A-171	Sequence 171, App
872	14	1.7	2309	2	US-08-449-609-10	Sequence 10, Appl	c 945	14	1.7	2847	4	US-09-556-877-186	Sequence 186, App
873	14	1.7	2330	5	PCT-US95-13975-71	Sequence 71, Appl	c 946	14	1.7	2847	4	US-09-620-412C-186	Sequence 186, App
874	14	1.7	2350	4	US-09-152-060-40	Sequence 40, Appl	c 947	14	1.7	2862	4	US-09-443-184-36	Sequence 36, Appl
875	14	1.7	2360	3	US-08-916-043-3	Sequence 3, Appl	c 948	14	1.7	2876	4	US-09-134-607A-9	Sequence 9, Appl
876	14	1.7	2364	4	US-09-172-045-1	Sequence 1, Appl	c 949	14	1.7	2876	4	US-09-134-607A-14	Sequence 14, Appl
877	14	1.7	2381	1	US-08-021-608D-9	Sequence 9, Appl	c 950	14	1.7	2897	4	US-09-134-607A-11	Sequence 11, Appl
878	14	1.7	2381	1	US-08-726-160-9	Sequence 9, Appl	c 951	14	1.7	2901	4	US-08-368-776A-4	Sequence 4, Appl
879	14	1.7	2381	5	PCT-US94-01782-9	Sequence 9, Appl	c 952	14	1.7	2901	5	PCT-US96-00419-4	Sequence 4, Appl
880	14	1.7	2383	4	US-09-192-909-1	Sequence 1, Appl	c 953	14	1.7	2912	1	US-07-931-943-1	Sequence 1, Appl
881	14	1.7	2384	1	US-08-021-608D-1	Sequence 1, Appl	c 954	14	1.7	2912	2	US-08-624-601-1	Sequence 1, Appl
882	14	1.7	2384	1	US-08-726-160-1	Sequence 1, Appl	c 955	14	1.7	2915	1	US-07-931-943-4	Sequence 4, Appl
883	14	1.7	2384	5	PCT-US94-01782-1	Sequence 1, Appl	c 956	14	1.7	2915	2	US-08-624-601-4	Sequence 4, Appl
884	14	1.7	2385	4	US-08-352-902D-145	Sequence 145, App	c 957	14	1.7	2940	2	US-08-418-848A-3	Sequence 3, Appl
885	14	1.7	2425	3	US-08-804-439A-5	Sequence 5, Appl	c 958	14	1.7	2945	6	5462872-1	Patent No. 5462872
886	14	1.7	2425	3	US-08-720-229-5	Sequence 5, Appl	c 959	14	1.7	2946	3	US-08-346-455B-35	Sequence 35, Appl
887	14	1.7	2426	3	US-08-028-327-2	Sequence 2, Appl	c 960	14	1.7	2946	3	US-08-977-221-35	Sequence 35, Appl
888	14	1.7	2467	2	US-08-701-240-1	Sequence 1, Appl	c 961	14	1.7	2946	3	US-09-483-831B-35	Sequence 35, Appl
889	14	1.7	2467	3	US-09-138-236-1	Sequence 1, Appl	c 962	14	1.7	2946	5	PCT-US95-06613-35	Sequence 35, Appl
890	14	1.7	2483	1	US-08-526-964-1	Sequence 1, Appl	c 963	14	1.7	2960	3	US-08-913-842-3	Sequence 3, Appl
891	14	1.7	2483	2	US-08-946-617-1	Sequence 2, Appl	c 964	14	1.7	2972	1	US-08-453-695A-114	Sequence 114, App
892	14	1.7	2483	3	US-09-031-897-1	Sequence 1, Appl	c 965	14	1.7	2972	1	US-08-268-161A-114	Sequence 114, App
893	14	1.7	2484	2	US-08-209-521-8	Sequence 8, Appl	c 966	14	1.7	2972	2	US-08-453-702A-114	Sequence 114, App
894	14	1.7	2484	4	US-08-961-810-4	Sequence 4, Appl	c 967	14	1.7	2972	2	US-09-099-639-114	Sequence 114, App
895	14	1.7	2484	4	US-08-352-902D-4	Sequence 4, Appl	c 968	14	1.7	2972	5	PCT-US95-08071-114	Sequence 114, App
896	14	1.7	2491	4	US-09-345-214-5	Sequence 5, Appl	c 969	14	1.7	2981	2	US-08-257-073-2	Sequence 2, Appl
897	14	1.7	2493	3	US-08-804-439A-11	Sequence 11, Appl	c 970	14	1.7	2981	2	US-08-184-009-119	Sequence 119, App
898	14	1.7	2493	3	US-08-720-229-11	Sequence 11, Appl	c 971	14	1.7	2981	2	US-08-458-356-119	Sequence 119, App
899	14	1.7	2508	4	US-09-198-122-7	Sequence 7, Appl	c 972	14	1.7	2981	2	US-08-460-736-119	Sequence 119, App
900	14	1.7	2508	2	US-08-850-993-1	Sequence 1, Appl	c 973	14	1.7	2990	1	US-08-572-951-1	Sequence 1, Appl
901	14	1.7	2515	3	US-08-714-918-37	Sequence 37, Appl	c 974	14	1.7	2995	4	US-09-149-476-280	Sequence 280, App
902	14	1.7	2515	4	US-09-265-315-37	Sequence 37, Appl	c 975	14	1.7	3018	4	US-09-206-942-40	Sequence 40, Appl
903	14	1.7	2515	4	US-09-265-315-37	Sequence 37, Appl	c 976	14	1.7				

```
14 1.7 3026 4 US-08-981-729-6 Sequence 6, Appli
14 1.7 3036 4 US-09-206-942-38 Sequence 38, Appli
14 1.7 3042 4 US-09-415-522-7 Sequence 7, Appli
14 1.7 3065 1 US-08-156-866-1 Sequence 1, Appli
14 1.7 3084 1 US-08-147-890-2 Sequence 2, Appli
14 1.7 3104 1 US-08-346-455B-66 Sequence 66, Appli
14 1.7 3104 3 US-08-977-221-66 Sequence 66, Appli
14 1.7 3104 4 US-09-483-831B-66 Sequence 66, Appli
14 1.7 3104 5 PCT-US95-06613-56 Sequence 66, Appli
14 1.7 3107 4 US-08-213-419B-1 Sequence 1, Appli
14 1.7 3112 4 US-08-961-527-116 Sequence 116, App
14 1.7 3114 4 US-09-134-001C-1957 Sequence 1957, Ap
14 1.7 3152 4 US-09-221-017B-931 Sequence 931, App
14 1.7 3167 4 US-09-276-531-132 Sequence 132, App
14 1.7 3182 1 US-07-797-556-5 Sequence 5, Appli
14 1.7 3182 1 US-07-943-843-1 Sequence 1, Appli
14 1.7 3182 1 US-08-347-003-1 Sequence 1, Appli
14 1.7 3213 2 US-08-975-527-2 Sequence 2, Appli
14 1.7 3216 2 US-08-828-007-1 Sequence 1, Appli
14 1.7 3231 4 US-09-134-001C-1200 Sequence 1200, Ap
14 1.7 3251 1 US-08-346-455B-68 Sequence 68, Appli
14 1.7 3251 3 US-08-977-221-68 Sequence 68, Appli
14 1.7 3251 4 US-09-483-831B-68 Sequence 68, Appli
14 1.7 3251 5 PCT-US95-06613-58 Sequence 68, Appli
```

## ALIGNMENTS

```
RESULT 1
US-09-660-587-41
; Sequence 41, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 41
; LENGTH: 840
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: nucleic acid sequence of E. canis p28-2
US-09-660-587-41
```

```
Query Match 100.0%; Score 840; DB 4; Length 840;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAATTATAGAAAATTTCTAGTAAGACGGGTTAATCTCAATTAATGTCAATCTTACCA 60
DB 1 ATGAATTATAGAAAATTTCTAGTAAGACGGGTTAATCTCAATTAATGTCAATCTTACCA 60

QY 61 TATCAGTCTTTTGCAGATCCTCTAGGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120
DB 61 TATCAGTCTTTTGCAGATCCTCTAGGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120

QY 121 ATTAGTCAAGATACAAATCCAACTATATCACACTTTAGAAAAATTTCTCTCGAAGAACT 180
DB 121 ATTAGTCAAGATACAAATCCAACTATATCACACTTTAGAAAAATTTCTCTCGAAGAACT 180

QY 181 CCTATTATGGAACAAATTTCTCTCAGTAAAGATTTTCGGACTAAAGAAAGATGGTAT 240
DB 181 CCTATTATGGAACAAATTTCTCTCAGTAAAGATTTTCGGACTAAAGAAAGATGGTAT 240
```

```
RESULT 2
US-09-660-587-37/c
; Sequence 37, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 37
; LENGTH: 22
; TYPE: DNA
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: primer Eca5'-1600
US-09-660-587-37
```

```
Query Match 2.6%; Score 22; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 670 GAAGTCTCTGCATTTATTGGTG 691
DB 22 GAAGTCTCTGCATTTATTGGTG 1
```



FACEIT NO. 3886780

COMPUTER READABLE FORM:

```
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/624,125
;; FILING DATE: 29-MAR-1996
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: KELBER, STEVEN B.
;; REGISTRATION NUMBER: 30,073
;; REFERENCE/DOCKET NUMBER: 2747-063-27
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 703-413-3000
;; TELEFAX: 703-413-2220
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 956 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
US-08-624-125-3

Query Match 2.1%; Score 18; DB 1; Length 956;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 808 TTTCGGCGGAGAAATTGGA 825
Db 203 TTTCGGCGGAGAAATTGGA 220
```

```
RESULT 7
US-07-638-431-1
; Sequence 1, Application US/07638431
; Patent No. 5198535
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenvit, Yupin
; APPLICANT: Hedstrom, Richard
; APPLICANT: Khumsmith, Srisin
; APPLICANT: Rogers IV, William O.
; TITLE OF INVENTION: Protective malaria sporozoite surface protein
; TITLE OF INVENTION: immunogen and gene
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: A. David Spevack
; STREET: NMDC Building 1 T-12 National Naval
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20814-5044
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/638,431
; FILING DATE: 19910110
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, Avrom D.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-4033
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4673 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
```

```
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: N
;; ANTI-SENSE: N
;; ORIGINAL SOURCE:
;; ORGANISM: Plasmodium yoelii
;; STRAIN: 17X(NL)
;; DEVELOPMENTAL STAGE: erythrocytic stage
;; TISSUE TYPE: Blood
;; CELL TYPE: erythrocytic stage
;; IMMEDIATE SOURCE:
;; LIBRARY: Py-lambda gt11-2-7 kb genomic expression
;; CLONE: Py10.1111
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 718..3195
;; OTHER INFORMATION:
US-07-638-431-1
```

```
Query Match 2.1%; Score 18; DB 1; Length 4673;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 312 TTCAGGAAGTATTGGTTA 329
Db 867 TTCAGGAAGTATTGGTTA 884
```

```
RESULT 8
PCT-US92-00018-1
; Sequence 1, Application PC/TUS9200018
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenvit, Yupin
; APPLICANT: Hedstrom, Richard
; APPLICANT: Khumsmith, Srisin
; APPLICANT: Rogers IV, William O.
; TITLE OF INVENTION: Protective malaria sporozoite surface protein
; TITLE OF INVENTION: immunogen and gene encoding
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: A. David Spevack
; STREET: NMDC Building 1 T-12 National Naval
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20814-5044
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00018
; FILING DATE: 19920103
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, Avram D.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-4033
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4673 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium yoelii
```

;; STRAIN: 17X(NL)  
;; DEVELOPMENTAL STAGE: erythrocytic stage  
;; TISSUE TYPE: Blood  
;; CELL TYPE: erythrocytic stage  
;; IMMEDIATE SOURCE:  
;; LIBRARY: Py-lambdagit11-2-7 kb genomic expression  
;; CLONE: Py10.1111  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 718..3195  
;; OTHER INFORMATION:  
PCT-US92-00018-1

Query Match 2.1%; Score 18; DB 5; Length 4673;  
Best Local Similarity 100.0%; Pred. No. 9.3;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 312 TTCAGGAAGTATTGGTTA 329  
Db 867 TTCAGGAAGTATTGGTTA 884

## RESULT 9

US-09-128-155-16  
; Sequence 16, Application US/09128155  
; Patent No. 6117654  
; GENERAL INFORMATION:  
; APPLICANT: Pan, Yang  
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY  
; FILE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: 09404/052001  
; CURRENT APPLICATION NUMBER: US/09/128,155  
; CURRENT FILING DATE: 1998-08-03  
; EARLIER APPLICATION NUMBER: US 60/091,650  
; EARLIER FILING DATE: 1998-07-02  
; EARLIER APPLICATION NUMBER: US 60/054,646  
; EARLIER FILING DATE: 1997-08-04  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 16  
; LENGTH: 152331  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(152331)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-128-155-16

Query Match 2.1%; Score 18; DB 3; Length 152331;  
Best Local Similarity 100.0%; Pred. No. 9.5;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 AATTATAGAAATTTCTA 21  
Db 61113 AATTATAGAAATTTCTA 61130

## RESULT 10

US-09-128-155-17/C  
; Sequence 17, Application US/09128155  
; Patent No. 6117654  
; GENERAL INFORMATION:  
; APPLICANT: Pan, Yang  
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY  
; FILE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: 09404/052001  
; CURRENT APPLICATION NUMBER: US/09/128,155  
; CURRENT FILING DATE: 1998-08-03  
; EARLIER APPLICATION NUMBER: US 60/091,650  
; EARLIER FILING DATE: 1998-07-02  
; EARLIER APPLICATION NUMBER: US 60/054,646  
; EARLIER FILING DATE: 1997-08-04

;; NUMBER OF SEQ ID NOS: 18  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 17  
;; LENGTH: 176373  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; LOCATION: (1)...(176373)  
;; OTHER INFORMATION: n = A,T,C or G  
US-09-128-155-17

Query Match 2.1%; Score 18; DB 3; Length 176373;  
Best Local Similarity 100.0%; Pred. No. 9.5;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 AATTATAGAAATTTCTA 21  
Db 110497 AATTATAGAAATTTCTA 110480

## RESULT 11

US-08-755-587-151/C  
; Sequence 151, Application US/08755587  
; Patent No. 6045997  
; GENERAL INFORMATION:  
; APPLICANT: Futreal, Phillip A  
; APPLICANT: Wooster, Richard F  
; APPLICANT: Ashworth, Alan  
; APPLICANT: Stratton, Michael R  
; TITLE OF INVENTION: Materials and methods relating to the  
; TITLE OF INVENTION: Identification and sequencing of the BRCA2 cancer  
; NUMBER OF SEQUENCES: 222  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Bell Seltzer Park & Gibson  
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107  
; CITY: Raleigh  
; STATE: NC  
; COUNTRY: USA  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/755,587  
; FILING DATE: 25-NOV-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9523959.6  
; FILING DATE: 23-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9525555.0  
; FILING DATE: 14-DEC-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9617961.9  
; FILING DATE: 28-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kenneth D Sibley  
; REGISTRATION NUMBER: 31,665  
; REFERENCE/DOCKET NUMBER: 5405-135  
; INFORMATION FOR SEQ ID NO: 151:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 24 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-755-587-151

Query Match 2.0%; Score 17; DB 3; Length 24;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 448 AAAGATGCAATGGAAGA 464  
DB 20 AAAGATGCAATGGAAGA 4

## RESULT 12

US-09-023-082A-35/c  
; Sequence 35, Application US/09023082A  
; Patent No. 6077692  
; GENERAL INFORMATION:  
; APPLICANT: RUBEN, STEVEN M.  
; APPLICANT: JIMENEZ, PABLO  
; APPLICANT: DUAN, D. ROXANNE  
; APPLICANT: RAMPY, MARK A.  
; APPLICANT: MENDRICK, DONNA  
; APPLICANT: ZHANG, JUN  
; APPLICANT: NI, JIAN  
; APPLICANT: MOORE, PAUL A.  
; APPLICANT: COLEMAN, TIMOTHY A.  
; APPLICANT: GRUBER, JOACHIM R.  
; APPLICANT: DILLON, PATRICK J.  
; APPLICANT: GENTZ, REINER L.  
; TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2  
; NUMBER OF SEQUENCES: 148  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
; STREET: 1100 NEW YORK AVE, NW, SUITE 600  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/023,082A  
; FILING DATE: 13-FEB-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/01790  
; FILING DATE: 14-FEB-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/461,195  
; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/023,852  
; FILING DATE: 13-AUG-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/039,045  
; FILING DATE: 28-FEB-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/862,432  
; FILING DATE: 23-MAY-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/910,875  
; FILING DATE: 13-AUG-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/055,561  
; FILING DATE: 13-AUG-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEFFFE, ERIC K.  
; REGISTRATION NUMBER: 1488.0360008/EKS  
; TELEPHONE: 202-371-2600  
; TELEFAX: 202-371-2540  
; INFORMATION FOR SEQ ID NO: 35:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 75 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single

; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
US-09-023-082A-35

Query Match 2.0%; Score 17; DB 3; Length 75;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 GATTTCAGGAAGTATT 324  
DB 45 GATTTCAGGAAGTATT 29

## RESULT 13

US-09-023-082A-103/c  
; Sequence 103, Application US/09023082A  
; Patent No. 6077692  
; GENERAL INFORMATION:  
; APPLICANT: RUBEN, STEVEN M.  
; APPLICANT: JIMENEZ, PABLO  
; APPLICANT: DUAN, D. ROXANNE  
; APPLICANT: RAMPY, MARK A.  
; APPLICANT: MENDRICK, DONNA  
; APPLICANT: ZHANG, JUN  
; APPLICANT: NI, JIAN  
; APPLICANT: MOORE, PAUL A.  
; APPLICANT: COLEMAN, TIMOTHY A.  
; APPLICANT: GRUBER, JOACHIM R.  
; APPLICANT: DILLON, PATRICK J.  
; APPLICANT: GENTZ, REINER L.  
; TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2  
; NUMBER OF SEQUENCES: 148  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
; STREET: 1100 NEW YORK AVE, NW, SUITE 600  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/023,082A  
; FILING DATE: 13-FEB-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/01790  
; FILING DATE: 14-FEB-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/461,195  
; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/023,852  
; FILING DATE: 13-AUG-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/039,045  
; FILING DATE: 28-FEB-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/862,432  
; FILING DATE: 23-MAY-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/910,875  
; FILING DATE: 13-AUG-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEFFFE, ERIC K.  
; REGISTRATION NUMBER: 1488.0360008/EKS  
; REFERENCE/DOCKET NUMBER: 1488.0360008/EKS

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 103:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 90 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-023-082A-103

Query Match          2.0%; Score 17; DB 3; Length 90;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 308 GATTTCAGGAAGTATT 324
Db 60 GATTTCAGGAAGTATT 44

RESULT 14
US-09-218-444-24/c
; Sequence 24, Application US/09218444
; Patent No. 6238888
; GENERAL INFORMATION:
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Chopra, Arvind
; APPLICANT: Kaushal, Parveen
; APPLICANT: Spitznagel, Thomas
; APPLICANT: Unsworth, Edward
; APPLICANT: Khan, Fazal
; TITLE OF INVENTION: Keratinocyte Growth Factor-2 Formulations
; FILE REFERENCE: 1488.1030001
; CURRENT APPLICATION NUMBER: US/09/218,444
; CURRENT FILING DATE: 1998-12-22
; EARLIER APPLICATION NUMBER: US 60/068,493
; EARLIER FILING DATE: 1997-12-22
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 90
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-218-444-24

Query Match          2.0%; Score 17; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 308 GATTTCAGGAAGTATT 324
Db 60 GATTTCAGGAAGTATT 44

RESULT 15
US-09-581-617-4/c
; Sequence 4, Application US/09581617
; Patent No. 6346385
; GENERAL INFORMATION:
; APPLICANT: Teljin Limited
; TITLE OF INVENTION: Analysis of predisposition based on human airway
; TITLE OF INVENTION: trypsin protease gene polymorphism
; FILE REFERENCE: 059572
; CURRENT APPLICATION NUMBER: US/09/581,617
; CURRENT FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: PCT/JP98/05689
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 9-346494 JAPAN
; PRIOR FILING DATE: 1997-12-16
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
```

```
; LENGTH: 177
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-581-617-4

Query Match          2.0%; Score 17; DB 4; Length 177;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 714 TAATAAAATTCAGGAAGA 730
Db 118 TAATAAAATTCAGGAAGA 102

RESULT 16
US-09-023-082A-81/c
; Sequence 81, Application US/09023082A
; Patent No. 607692
; GENERAL INFORMATION:
; APPLICANT: RUBEN, STEVEN M.
; APPLICANT: JIMENEZ, PABLO
; APPLICANT: DUAN, D. ROXANNE
; APPLICANT: RAMPY, MARK A.
; APPLICANT: MENDRICK, DONNA
; APPLICANT: ZHANG, JUN
; APPLICANT: NI, JIAN
; APPLICANT: MOORE, PAUL A.
; APPLICANT: COLEMAN, TIMOTHY A.
; APPLICANT: GRUBER, JOACHIM R.
; APPLICANT: DILLON, PATRICK J.
; APPLICANT: GENTZ, REINER L.
; TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,082A
; FILING DATE: 13-FEB-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/01790
; FILING DATE: 14-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/461,195
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/023,852
; FILING DATE: 13-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/039,045
; FILING DATE: 28-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,432
; FILING DATE: 23-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/910,875
; FILING DATE: 13-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,561
; FILING DATE: 13-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFFE, ERIC K.
; REGISTRATION NUMBER: 36,688.
```

REFERENCE/DOCKET NUMBER: 1488.0360008/EKS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2500  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 81:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 276 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: both  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..276  
US-09-023-082A-81

Query Match 2.0%; Score 17; DB 3; Length 276;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 GATTTCAGGAAGTATT 324  
|||||  
DB 93 GATTTCAGGAAGTATT 77

RESULT 17  
US-09-023-082A-79/c  
Sequence 79, Application US/09023082A  
Patent No. 6077692  
GENERAL INFORMATION:  
APPLICANT: RUBEN, STEVEN M.  
APPLICANT: JIMENEZ, PABLO  
APPLICANT: DUAN, D. ROXANNE  
APPLICANT: RAMPY, MARK A.  
APPLICANT: MENDRICK, DONNA  
APPLICANT: ZHANG, JUN  
APPLICANT: NI, JIAN  
APPLICANT: MOORE, PAUL A.  
APPLICANT: COLEMAN, TIMOTHY A.  
APPLICANT: GRUBER, JOACHIM R.  
APPLICANT: DILLON, PATRICK J.  
APPLICANT: GENTZ, REINER L.  
TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2  
NUMBER OF SEQUENCES: 148  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
STREET: 1100 NEW YORK AVE, NW, SUITE 600  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION NUMBER: US/09/023,082A  
FILING DATE: 13-FEB-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/01790  
FILING DATE: 14-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/461,195  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/023,852  
FILING DATE: 13-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/039,045  
FILING DATE: 28-FEB-1997  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/862,432  
FILING DATE: 23-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/910,875  
FILING DATE: 13-AUG-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/055,561  
FILING DATE: 13-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: STEFFFE, ERIC K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.0360008/EKS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 79:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 357 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: both  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..357  
US-09-023-082A-79

Query Match 2.0%; Score 17; DB 3; Length 357;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 GATTTCAGGAAGTATT 324  
|||||  
DB 174 GATTTCAGGAAGTATT 158

RESULT 18  
US-09-023-082A-69/c  
Sequence 69, Application US/09023082A  
Patent No. 6077692  
GENERAL INFORMATION:  
APPLICANT: RUBEN, STEVEN M.  
APPLICANT: JIMENEZ, PABLO  
APPLICANT: DUAN, D. ROXANNE  
APPLICANT: RAMPY, MARK A.  
APPLICANT: MENDRICK, DONNA  
APPLICANT: ZHANG, JUN  
APPLICANT: NI, JIAN  
APPLICANT: MOORE, PAUL A.  
APPLICANT: COLEMAN, TIMOTHY A.  
APPLICANT: GRUBER, JOACHIM R.  
APPLICANT: DILLON, PATRICK J.  
APPLICANT: GENTZ, REINER L.  
TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2  
NUMBER OF SEQUENCES: 148  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
STREET: 1100 NEW YORK AVE, NW, SUITE 600  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,082A  
FILING DATE: 13-FEB-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/01790

;; FILING DATE: 14-FEB-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/461,195  
;; FILING DATE: 05-JUN-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 60/023,852  
;; FILING DATE: 13-AUG-1996  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 60/039,045  
;; FILING DATE: 28-FEB-1997  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/862,432  
;; FILING DATE: 23-MAY-1997  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/910,875  
;; FILING DATE: 13-AUG-1997  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 60/055,561  
;; FILING DATE: 13-AUG-1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: STEFFFE, ERIC K.  
;; REGISTRATION NUMBER: 36,688  
;; REFERENCE/DOCKET NUMBER: 1488.0360008/EKS  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-371-2600  
;; TELEFAX: 202-371-2540  
;; INFORMATION FOR SEQ ID NO: 69:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 402 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: both  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 1..399  
US-09-023-082A-69

Query Match 2.0%; Score 17; DB 3; Length 402;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 308 GATTTCAGGAAGTATT 324  
Db 51 GATTTCAGGAAGTATT 35

RESULT 19  
US-09-023-082A-95/c  
; Sequence 95, Application US/09023082A  
; Patent No. 6077692  
; GENERAL INFORMATION:  
; APPLICANT: RUBEN, STEVEN M.  
; APPLICANT: JIMENEZ, PABLO  
; APPLICANT: DUAN, D. ROXANNE  
; APPLICANT: RAMPY, MARK A.  
; APPLICANT: MENDRICK, DONNA  
; APPLICANT: ZHANG, JUN  
; APPLICANT: NI, JIAN  
; APPLICANT: MOORE, PAUL A.  
; APPLICANT: COLEMAN, TIMOTHY A.  
; APPLICANT: GRUBER, JOACHIM R.  
; APPLICANT: DILLON, PATRICK J.  
; APPLICANT: GENTZ, REINER L.  
; TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2  
; NUMBER OF SEQUENCES: 148  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
; STREET: 1100 NEW YORK AVE, NW, SUITE 600  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3934

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA: US/09/023,082A  
;; APPLICATION NUMBER: US 08/910,875  
;; FILING DATE: 13-FEB-1998  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US95/01790  
;; FILING DATE: 14-FEB-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/461,195  
;; FILING DATE: 05-JUN-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 60/023,852  
;; FILING DATE: 13-AUG-1996  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 60/039,045  
;; FILING DATE: 28-FEB-1997  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/862,432  
;; FILING DATE: 23-MAY-1997  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/910,875  
;; FILING DATE: 13-AUG-1997  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 60/055,561  
;; FILING DATE: 13-AUG-1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: STEFFFE, ERIC K.  
;; REGISTRATION NUMBER: 36,688  
;; REFERENCE/DOCKET NUMBER: 1488.0360008/EKS  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-371-2600  
;; TELEFAX: 202-371-2540  
;; INFORMATION FOR SEQ ID NO: 95:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 426 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cDNA  
US-09-023-082A-95

Query Match 2.0%; Score 17; DB 3; Length 426;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 308 GATTTCAGGAAGTATT 324  
Db 75 GATTTCAGGAAGTATT 59

RESULT 20  
US-09-023-082A-111/c  
; Sequence 111, Application US/09023082A  
; Patent No. 6077692  
; GENERAL INFORMATION:  
; APPLICANT: RUBEN, STEVEN M.  
; APPLICANT: JIMENEZ, PABLO  
; APPLICANT: DUAN, D. ROXANNE  
; APPLICANT: RAMPY, MARK A.  
; APPLICANT: MENDRICK, DONNA  
; APPLICANT: ZHANG, JUN  
; APPLICANT: NI, JIAN  
; APPLICANT: MOORE, PAUL A.  
; APPLICANT: COLEMAN, TIMOTHY A.  
; APPLICANT: GRUBER, JOACHIM R.  
; APPLICANT: DILLON, PATRICK J.  
; APPLICANT: GENTZ, REINER L.  
; TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2

NUMBER OF SEQUENCES: 148  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
STREET: 1100 NEW YORK AVE, NW, SUITE 600  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,082A  
FILING DATE: 13-FEB-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION NUMBER: PCT/US95/01790  
FILING DATE: 14-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/461,195  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/023,852  
FILING DATE: 13-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/039,045  
FILING DATE: 28-FEB-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/862,432  
FILING DATE: 23-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/055,561  
FILING DATE: 13-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: STEFFE, ERIC K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.0360008/EKS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2540  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 111:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 426 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-09-023-082A-111

Query Match 2.0%; Score 17; DB 3; Length 426;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 GATTTCAGGAAGTATT 324  
|||||  
Db 75 GATTTCAGGAAGTATT 59

RESULT 21  
US-09-023-082A-119/c  
Sequence 119, Application US/09023082A  
Patent No. 6077692  
GENERAL INFORMATION:  
APPLICANT: RUBEN, STEVEN M.  
APPLICANT: JIMENEZ, PABLO  
APPLICANT: DUAN, D. ROXANNE  
APPLICANT: RAMPY, MARK A.  
APPLICANT: MENDRICK, DONNA

APPLICANT: ZHANG, JUN  
APPLICANT: NI, JIAN  
APPLICANT: MOORE, PAUL A.  
APPLICANT: COLEMAN, TIMOTHY A.  
APPLICANT: GRUBER, JOACHIM R.  
APPLICANT: DILLON, PATRICK J.  
APPLICANT: GENTZ, REINER L.  
TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2  
NUMBER OF SEQUENCES: 148  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
STREET: 1100 NEW YORK AVE, NW, SUITE 600  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,082A  
FILING DATE: 13-FEB-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/01790  
FILING DATE: 14-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/461,195  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/023,852  
FILING DATE: 13-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/039,045  
FILING DATE: 28-FEB-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/862,432  
FILING DATE: 23-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/910,875  
FILING DATE: 13-AUG-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/055,561  
FILING DATE: 13-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: STEFFE, ERIC K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.0360008/EKS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2540  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 119:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 426 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-09-023-082A-119

Query Match 2.0%; Score 17; DB 3; Length 426;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 GATTTCAGGAAGTATT 324  
|||||  
Db 75 GATTTCAGGAAGTATT 59

RESULT 22  
US-09-023-082A-123/c



Sequence 123, Application US/09023082A  
Patent No. 6077692  
GENERAL INFORMATION:  
APPLICANT: RUBEN, STEVEN M.  
APPLICANT: JIMENEZ, PABLO  
APPLICANT: DUAN, D. ROXANNE  
APPLICANT: RAMPY, MARK A.  
APPLICANT: MENDRICK, DONNA  
APPLICANT: ZHANG, JUN  
APPLICANT: NI, JIAN  
APPLICANT: MOORE, PAUL A.  
APPLICANT: COLEMAN, TIMOTHY A.  
APPLICANT: GRUBER, JOACHIM R.  
APPLICANT: DILLON, PATRICK J.  
APPLICANT: GENTZ, REINER L.  
TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2  
NUMBER OF SEQUENCES: 148  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
STREET: 1100 NEW YORK AVE, NW, SUITE 600  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,082A  
FILING DATE: 13-FEB-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/01790  
FILING DATE: 14-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/461,195  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/023,852  
FILING DATE: 13-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/039,045  
FILING DATE: 28-FEB-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/862,432  
FILING DATE: 23-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/055,561  
FILING DATE: 13-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: STEFFE, ERIC K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.0360008/EKS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 123:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 426 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
US-09-023-082A-123

Query Match 2.0%; Score 17; DB 3; Length 426;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 308 GATTTTCAGGAAGTATT 324  
|||||  
Db 75 GATTTTCAGGAAGTATT 59

RESULT 23  
US-09-023-082A-127/c  
Sequence 127, Application US/09023082A  
Patent No. 6077692  
GENERAL INFORMATION:  
APPLICANT: RUBEN, STEVEN M.  
APPLICANT: JIMENEZ, PABLO  
APPLICANT: DUAN, D. ROXANNE  
APPLICANT: RAMPY, MARK A.  
APPLICANT: MENDRICK, DONNA  
APPLICANT: ZHANG, JUN  
APPLICANT: NI, JIAN  
APPLICANT: MOORE, PAUL A.  
APPLICANT: COLEMAN, TIMOTHY A.  
APPLICANT: GRUBER, JOACHIM R.  
APPLICANT: DILLON, PATRICK J.  
APPLICANT: GENTZ, REINER L.  
TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2  
NUMBER OF SEQUENCES: 148  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
STREET: 1100 NEW YORK AVE, NW, SUITE 600  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,082A  
FILING DATE: 13-FEB-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/01790  
FILING DATE: 14-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/461,195  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/023,852  
FILING DATE: 13-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/039,045  
FILING DATE: 28-FEB-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/862,432  
FILING DATE: 23-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/910,875  
FILING DATE: 13-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: STEFFE, ERIC K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.0360008/EKS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 127:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 426 base pairs  
TYPE: nucleic acid

```
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-09-023-082A-127

Query Match          2.0%; Score 17; DB 3; Length 426;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 GATTTTCAGGAAGTATT 324
Db 75 GATTTTCAGGAAGTATT 59

RESULT 24
US-09-023-082A-131/c
; Sequence 131, Application US/09023082A
; Patent No. 6077692
; GENERAL INFORMATION:
; APPLICANT: RUBEN, STEVEN M.
; APPLICANT: JIMENEZ, PABLO
; APPLICANT: DUAN, D. ROXANNE
; APPLICANT: RAMPY, MARK A.
; APPLICANT: MENDRICK, DONNA
; APPLICANT: ZHANG, JUN
; APPLICANT: NI, JIAN
; APPLICANT: MOORE, PAUL A.
; APPLICANT: COLEMAN, TIMOTHY A.
; APPLICANT: GRUBER, JOACHIM R.
; APPLICANT: DILLON, PATRICK J.
; APPLICANT: GENTZ, REINER L.
; TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,082A
; FILING DATE: 13-FEB-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/01790
; FILING DATE: 14-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/461,195
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/023,852
; FILING DATE: 13-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/039,045
; FILING DATE: 28-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,432
; FILING DATE: 23-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/910,875
; FILING DATE: 13-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,561
; FILING DATE: 13-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFFE, ERIC K.
; REGISTRATION NUMBER: 36,688

REFERENCE/DOCKET NUMBER: 1488.0360008/EKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 131:
SEQUENCE CHARACTERISTICS:
LENGTH: 426 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-023-082A-131

Query Match          2.0%; Score 17; DB 3; Length 426;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 GATTTTCAGGAAGTATT 324
Db 75 GATTTTCAGGAAGTATT 59

RESULT 25
US-09-023-082A-135/c
; Sequence 135, Application US/09023082A
; Patent No. 6077692
; GENERAL INFORMATION:
; APPLICANT: RUBEN, STEVEN M.
; APPLICANT: JIMENEZ, PABLO
; APPLICANT: DUAN, D. ROXANNE
; APPLICANT: RAMPY, MARK A.
; APPLICANT: MENDRICK, DONNA
; APPLICANT: ZHANG, JUN
; APPLICANT: NI, JIAN
; APPLICANT: MOORE, PAUL A.
; APPLICANT: COLEMAN, TIMOTHY A.
; APPLICANT: GRUBER, JOACHIM R.
; APPLICANT: DILLON, PATRICK J.
; APPLICANT: GENTZ, REINER L.
; TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,082A
; FILING DATE: 13-FEB-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/01790
; FILING DATE: 14-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/461,195
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/023,852
; FILING DATE: 13-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/039,045
; FILING DATE: 28-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,432
; FILING DATE: 23-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/910,875
; FILING DATE: 13-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,561
; FILING DATE: 13-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
```

APPLICATION NUMBER: US 08/910,875  
FILING DATE: 13-AUG-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/055,561  
FILING DATE: 13-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: STEFFFE, ERIC K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.0360008/EKS  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 135:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 426 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
US-09-023-082A-135

Query Match 2.0%; Score 17; DB 3; Length 426;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 308 GATTTCAGGAAGTATT 324  
Db 75 GATTTCAGGAAGTATT 59

RESULT 26  
US-09-023-082A-139/c  
Sequence 139, Application US/09023082A  
Patent No. 6077692  
GENERAL INFORMATION:  
APPLICANT: RUBEN, STEVEN M.  
APPLICANT: JIMENEZ, PABLO  
APPLICANT: DUAN, D. ROXANNE  
APPLICANT: RAMPY, MARK A.  
APPLICANT: MENDRICK, DONNA  
APPLICANT: ZHANG, JUN  
APPLICANT: NI, JIAN  
APPLICANT: MOORE, PAUL A.  
APPLICANT: COLEMAN, TIMOTHY A.  
APPLICANT: GRUBER, JOACHIM R.  
APPLICANT: DILLON, PATRICK J.  
APPLICANT: GENTZ, REINER L.  
TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2  
NUMBER OF SEQUENCES: 148  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
STREET: 1100 NEW YORK AVE, NW, SUITE 600  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023.082A  
FILING DATE: 13-FEB-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/01790  
FILING DATE: 14-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/461,195  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/023,852

FILING DATE: 13-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/039,045  
FILING DATE: 28-FEB-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/862,432  
FILING DATE: 23-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/910,875  
FILING DATE: 13-AUG-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/055,561  
FILING DATE: 13-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: STEFFFE, ERIC K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.0360008/EKS  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 139:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 426 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
US-09-023-082A-139

Query Match 2.0%; Score 17; DB 3; Length 426;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 308 GATTTCAGGAAGTATT 324  
Db 75 GATTTCAGGAAGTATT 59

RESULT 27  
US-09-023-082A-145/c  
Sequence 145, Application US/09023082A  
Patent No. 6077692  
GENERAL INFORMATION:  
APPLICANT: RUBEN, STEVEN M.  
APPLICANT: JIMENEZ, PABLO  
APPLICANT: DUAN, D. ROXANNE  
APPLICANT: RAMPY, MARK A.  
APPLICANT: MENDRICK, DONNA  
APPLICANT: ZHANG, JUN  
APPLICANT: NI, JIAN  
APPLICANT: MOORE, PAUL A.  
APPLICANT: COLEMAN, TIMOTHY A.  
APPLICANT: GRUBER, JOACHIM R.  
APPLICANT: DILLON, PATRICK J.  
APPLICANT: GENTZ, REINER L.  
TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2  
NUMBER OF SEQUENCES: 148  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
STREET: 1100 NEW YORK AVE, NW, SUITE 600  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,082A  
FILING DATE: 13-FEB-1998  
CLASSIFICATION: 435

```

RESULT 28
US-09-218-444-16/c
; Sequence 16, Application US/09218444
; Patent No. 6238888
; GENERAL INFORMATION:
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Chopra, Arvind
; APPLICANT: Kaushal, Parveen
; APPLICANT: Spitznagel, Thomas
; APPLICANT: Unsworth, Edward
; APPLICANT: Khan, Fazal
; TITLE OF INVENTION: Keratinocyte Growth Factor-2 Formulations
; FILE REFERENCE: 1488.1030001
; CURRENT APPLICATION NUMBER: US/09/218,444
; CURRENT FILING DATE: 1998-12-22
; EARLIER APPLICATION NUMBER: US 60/068,493
; EARLIER FILING DATE: 1997-12-22
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 426
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-218-444-16

```

APPLICATION NUMBER: US/09/023,082A  
FILING DATE: 13-FEB-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/01790  
FILING DATE: 14-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/461,195  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/023,852  
FILING DATE: 13-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/039,045  
FILING DATE: 28-FEB-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/862,432  
FILING DATE: 23-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/910,875  
FILING DATE: 13-AUG-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/055,561  
FILING DATE: 13-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: STEFFFE, ERIC K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.0360008/EKS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 67:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 444 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: both  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..441  
US-09-023-082A-67

Query Match 2.0%; Score 17; DB 3; Length 444;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 308 GATTTTCAGGAAGTATT 324  
Db 93 GATTTTCAGGAAGTATT 77

RESULT 31  
US-09-023-082A-115/c  
Sequence 115, Application US/09023082A  
Patent No. 6077692  
GENERAL INFORMATION:  
APPLICANT: RUBEN, STEVEN M.  
APPLICANT: JIMENEZ, PABLO  
APPLICANT: DUAN, D. ROXANNE  
APPLICANT: RAMPY, MARK A.  
APPLICANT: MENDRICK, DONNA  
APPLICANT: ZHANG, JUN  
APPLICANT: NI, JIAN  
APPLICANT: MOORE, PAUL A.  
APPLICANT: COLEMAN, TIMOTHY A.  
APPLICANT: GRUBER, JOACHIM R.  
APPLICANT: DILLON, PATRICK J.  
APPLICANT: GENTZ, REINER L.  
TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2  
NUMBER OF SEQUENCES: 148  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNER, KESSLER, GOLDSTEIN & FOX, P.L.L.C.

STREET: 1100 NEW YORK AVE, NW, SUITE 600  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,082A  
FILING DATE: 13-FEB-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/01790  
FILING DATE: 14-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/461,195  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/023,852  
FILING DATE: 13-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/039,045  
FILING DATE: 28-FEB-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/862,432  
FILING DATE: 23-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/910,875  
FILING DATE: 13-AUG-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/055,561  
FILING DATE: 13-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: STEFFFE, ERIC K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.0360008/EKS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 115:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 516 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-09-023-082A-115

Query Match 2.0%; Score 17; DB 3; Length 516;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 308 GATTTTCAGGAAGTATT 324  
Db 165 GATTTTCAGGAAGTATT 149

RESULT 32  
US-09-023-082A-42/c  
Sequence 42, Application US/09023082A  
Patent No. 6077692  
GENERAL INFORMATION:  
APPLICANT: RUBEN, STEVEN M.  
APPLICANT: JIMENEZ, PABLO  
APPLICANT: DUAN, D. ROXANNE  
APPLICANT: RAMPY, MARK A.  
APPLICANT: MENDRICK, DONNA  
APPLICANT: ZHANG, JUN  
APPLICANT: NI, JIAN  
APPLICANT: MOORE, PAUL A.

APPLICANT: COLEMAN, TIMOTHY A.  
APPLICANT: GRUBER, JOACHIM R.  
APPLICANT: DILLON, PATRICK J.  
APPLICANT: GENTZ, REINER L.  
TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2  
NUMBER OF SEQUENCES: 148  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
STREET: 1100 NEW YORK AVE, NW, SUITE 600  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,082A  
FILING DATE: 13-FEB-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/01790  
FILING DATE: 14-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/461,195  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/039,045  
FILING DATE: 28-FEB-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/862,432  
FILING DATE: 23-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/055,561  
FILING DATE: 13-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: STEFFER, ERIC K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.0360008/EKS  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 525 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: Both  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..522  
US-09-023-082A-42

Query Match 2.0%; Score 17; DB 3; Length 525;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 GATTTCAGGAAGTATT 324

Db 174 GATTTCAGGAAGTATT 158

RESULT 33

US-09-023-082A-65/c

Sequence 65, Application US/09023082A  
Patent No. 6077892  
GENERAL INFORMATION:  
APPLICANT: RUBEN, STEVEN M.  
APPLICANT: JIMENEZ, PABLO  
APPLICANT: DUAN, D. ROXANNE  
APPLICANT: RAMPY, MARK A.  
APPLICANT: MENDRICK, DONNA  
APPLICANT: ZHANG, JUN  
APPLICANT: NI, JIAN  
APPLICANT: MOORE, PAUL A.  
APPLICANT: COLEMAN, TIMOTHY A.  
APPLICANT: GRUBER, JOACHIM R.  
APPLICANT: DILLON, PATRICK J.  
APPLICANT: GENTZ, REINER L.  
TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2  
NUMBER OF SEQUENCES: 148  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
STREET: 1100 NEW YORK AVE, NW, SUITE 600  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,082A  
FILING DATE: 13-FEB-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/01790  
FILING DATE: 14-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/461,195  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/023,852  
FILING DATE: 13-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/039,045  
FILING DATE: 28-FEB-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/862,432  
FILING DATE: 23-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/910,875  
FILING DATE: 13-AUG-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/055,561  
FILING DATE: 13-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: STEFFER, ERIC K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.0360008/EKS  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 525 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: Both  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..522  
US-09-023-082A-65



```
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0360008/EKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 525 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: both
; MOLECULE TYPE: DNA (genomic)
; US-09-023-082A-84
;
Query Match          2.0%; Score 17; DB 3; Length 525;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 308 GATTTTCAGGAAGTATT 324
   |||||
Db 174 GATTTTCAGGAAGTATT 158
;
RESULT 36
US-09-023-082A-29/c
; Sequence 29, Application US/09023082A
; Patent No. 6077692
; GENERAL INFORMATION:
; APPLICANT: RUBEN, STEVEN M.
; APPLICANT: JIMENEZ, PABLO
; APPLICANT: DUAN, D. ROXANNE
; APPLICANT: RAMPY, MARK A.
; APPLICANT: MENDRICK, DONNA
; APPLICANT: ZHANG, JUN
; APPLICANT: NI, JIAN
; APPLICANT: MOORE, PAUL A.
; APPLICANT: COLEMAN, TIMOTHY A.
; APPLICANT: GRUBER, JOACHIM R.
; APPLICANT: DILLON, PATRICK J.
; APPLICANT: GENTZ, REINER L.
; TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,082A
; FILING DATE: 13-FEB-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/01790
; FILING DATE: 14-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/461,195
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/023,852
; FILING DATE: 13-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/039,045
; FILING DATE: 28-FEB-1997
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: US 08/862,432
; FILING DATE: 23-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/910,875
; FILING DATE: 13-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,561
; FILING DATE: 13-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0360008/EKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 555 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..553
; US-09-023-082A-29
;
Query Match          2.0%; Score 17; DB 3; Length 555;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 308 GATTTTCAGGAAGTATT 324
   |||||
Db 204 GATTTTCAGGAAGTATT 189
;
RESULT 37
US-09-023-082A-38/c
; Sequence 38, Application US/09023082A
; Patent No. 6077692
; GENERAL INFORMATION:
; APPLICANT: RUBEN, STEVEN M.
; APPLICANT: JIMENEZ, PABLO
; APPLICANT: DUAN, D. ROXANNE
; APPLICANT: RAMPY, MARK A.
; APPLICANT: MENDRICK, DONNA
; APPLICANT: ZHANG, JUN
; APPLICANT: NI, JIAN
; APPLICANT: MOORE, PAUL A.
; APPLICANT: COLEMAN, TIMOTHY A.
; APPLICANT: GRUBER, JOACHIM R.
; APPLICANT: DILLON, PATRICK J.
; APPLICANT: GENTZ, REINER L.
; TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,082A
; FILING DATE: 13-FEB-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/01790
```



;; FILING DATE: 14-FEB-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/461,195  
;; FILING DATE: 05-JUN-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 60/023,852  
;; FILING DATE: 13-AUG-1996  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 60/039,045  
;; FILING DATE: 28-FEB-1997  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/862,432  
;; FILING DATE: 23-MAY-1997  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/910,875  
;; FILING DATE: 13-AUG-1997  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 60/055,561  
;; FILING DATE: 13-AUG-1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: STEFFEE, ERIC K.  
;; REGISTRATION NUMBER: 36,688  
;; REFERENCE/DOCKET NUMBER: 1488.0360008/EKS  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-371-2600  
;; TELEFAX: 202-371-2540  
;; INFORMATION FOR SEQ ID NO: 38:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 627 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: both  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 1..624  
US-09-023-082A-38

Query Match 2.0%; Score 17; DB 3; Length 627;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 308 GATTTTCAGGAAGTATT 324  
Db 276 GATTTTCAGGAAGTATT 260

RESULT 38  
US-08-755-587-40/c  
; Sequence 40, Application US/08755587  
; Patent No. 6045997  
; GENERAL INFORMATION:  
; APPLICANT: Futreal, Phillip A  
; APPLICANT: Wooster, Richard F  
; APPLICANT: Ashworth, Alan  
; APPLICANT: Stratton, Michael R  
; TITLE OF INVENTION: Materials and methods relating to the  
; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer  
; TITLE OF INVENTION: susceptibility gene and uses thereof.  
; NUMBER OF SEQUENCES: 222  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Bell Seltzer Park & Gibson  
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107  
; CITY: Raleigh  
; STATE: NC  
; COUNTRY: USA  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/755,587

;; FILING DATE: 25-NOV-1996  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: GB 9523959.6  
;; FILING DATE: 23-NOV-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: GB 9525555.0  
;; FILING DATE: 14-DEC-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: GB 9617961.9  
;; FILING DATE: 28-AUG-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Kenneth D Sibley  
;; REGISTRATION NUMBER: 31,665  
;; REFERENCE/DOCKET NUMBER: 5405-135  
;; INFORMATION FOR SEQ ID NO: 40:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1139 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: exon  
;; LOCATION: 501..639  
;; NAME/KEY: CDS  
;; LOCATION: 501..639  
US-08-755-587-40

Query Match 2.0%; Score 17; DB 3; Length 1139;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 448 AAAGATGCAATGGAAGA 464  
Db 224 AAAGATGCAATGGAAGA 208

RESULT 39  
US-08-755-587-39/c  
; Sequence 39, Application US/08755587  
; Patent No. 6045997  
; GENERAL INFORMATION:  
; APPLICANT: Futreal, Phillip A  
; APPLICANT: Wooster, Richard F  
; APPLICANT: Ashworth, Alan  
; APPLICANT: Stratton, Michael R  
; TITLE OF INVENTION: Materials and methods relating to the  
; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer  
; TITLE OF INVENTION: susceptibility gene and uses thereof.  
; NUMBER OF SEQUENCES: 222  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Bell Seltzer Park & Gibson  
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107  
; CITY: Raleigh  
; STATE: NC  
; COUNTRY: USA  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/755,587  
; FILING DATE: 25-NOV-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9523959.6  
; FILING DATE: 23-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9525555.0  
; FILING DATE: 14-DEC-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9617961.9

;; FILING DATE: 28-AUG-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Kenneth D Sibley  
;; REGISTRATION NUMBER: 31,665  
;; REFERENCE/DOCKET NUMBER: 5405-135  
;; INFORMATION FOR SEQ ID NO: 39:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1164 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: exon  
;; LOCATION: 501..664  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 503..664  
US-08-755-587-39

Query Match 2.0%; Score 17; DB 3; Length 1164;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 448 AAAGATGCAATGGAAGA 464  
Db 481 AAAGATGCAATGGAAGA 465

## RESULT 40

US-08-755-587-38/c  
; Sequence 38, Application US/08755587  
; Patent No. 6045997  
; GENERAL INFORMATION:  
; APPLICANT: Futreal, Phillip A  
; APPLICANT: Wooster, Richard P  
; APPLICANT: Ashworth, Alan  
; APPLICANT: Stratton, Michael R  
; TITLE OF INVENTION: Materials and methods relating to the  
; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer  
; NUMBER OF SEQUENCES: 222  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Bell Seltzer Park & Gibson  
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107  
; CITY: Raleigh  
; STATE: NC  
; COUNTRY: USA  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/755,587  
; FILING DATE: 25-NOV-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9523959.6  
; FILING DATE: 23-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9525555.0  
; FILING DATE: 14-DEC-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9617961.9  
; FILING DATE: 28-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kenneth D Sibley  
; REGISTRATION NUMBER: 31,665  
; REFERENCE/DOCKET NUMBER: 5405-135  
; INFORMATION FOR SEQ ID NO: 38:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1199 base pairs  
; TYPE: nucleic acid

;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: exon  
;; LOCATION: 501..699  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 501..699  
US-08-755-587-38

Query Match 2.0%; Score 17; DB 3; Length 1199;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 448 AAAGATGCAATGGAAGA 464  
Db 914 AAAGATGCAATGGAAGA 898

Search completed: July 4, 2003, 02:34:15  
Job time : 68 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 02:15:42 ; Search time 131 Seconds  
(without alignments)  
10009.411 Million cell updates/sec

Title: US-10-062-624-41  
Perfect score: 840  
Sequence: 1 atgaattataagaaattctt.....ttggaatgaggttcaccttc 840

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1085931 seqs, 780495707 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Match	Length	ID	Description
1	840	100.0	840 9	US-10-062-051-41
2	840	100.0	840 9	US-10-062-920-41
3	840	100.0	840 12	US-10-062-624-41
4	840	100.0	843 9	US-10-314-639-47
5	840	100.0	843 12	US-10-059-964-47
6	30	3.6	852 9	US-10-314-639-3
7	30	3.6	852 12	US-10-059-964-3
8	22	2.6	22 9	US-10-062-051-37
9	22	2.6	22 9	US-10-062-920-37
10	22	2.6	22 12	US-10-062-624-37
11	19	2.3	840 9	US-10-062-051-5
12	19	2.3	840 9	US-10-062-920-5
13	19	2.3	840 12	US-10-062-624-5
14	19	2.3	843 9	US-10-314-639-37
15	19	2.3	843 12	US-10-059-964-37
16	19	2.3	894 9	US-10-314-639-13
17	19	2.3	894 12	US-10-059-964-13
18	19	2.3	5279 9	US-10-239-676-18
19	18	2.1	4929 10	US-09-864-761-1982

20	18	2.1	495	9	US-09-770-961-540	Sequence 540, App
21	18	2.1	588	10	US-09-864-761-8939	Sequence 8939, App
22	18	2.1	933	9	US-09-938-842A-1696	Sequence 1696, App
23	18	2.1	956	10	US-09-323-998D-3	Sequence 3, Appli
24	18	2.1	1301	12	US-10-001-843-25	Sequence 25, Appli
25	18	2.1	4923	7	US-08-781-986A-551	Sequence 551, App
26	18	2.1	6617	9	US-10-210-296-1	Sequence 1, Appli
27	18	2.1	152331	9	US-10-095-407-16	Sequence 16, Appli
28	18	2.1	170834	10	US-09-835-233-7	Sequence 17, Appli
29	18	2.1	176373	9	US-10-095-407-17	Sequence 17, Appli
30	18	2.1	183337	9	US-10-020-141-5	Sequence 5, Appli
31	18	2.1	368004	10	US-09-949-654-3	Sequence 3, Appli
32	17	2.0	75	9	US-09-345-373-35	Sequence 35, Appli
33	17	2.0	75	9	US-10-194-443-19	Sequence 19, Appli
34	17	2.0	90	9	US-09-345-373-103	Sequence 103, App
35	17	2.0	90	10	US-09-853-666-24	Sequence 24, Appli
36	17	2.0	256	10	US-09-878-574-10764	Sequence 10764, A
37	17	2.0	276	9	US-09-345-373-81	Sequence 81, Appli
38	17	2.0	296	9	US-10-091-572-579	Sequence 579, App
39	17	2.0	296	9	US-10-091-572-580	Sequence 580, App
40	17	2.0	296	9	US-09-764-891-6723	Sequence 6723, Ap
41	17	2.0	296	9	US-09-764-891-6724	Sequence 6724, Ap
42	17	2.0	307	9	US-10-046-935-710	Sequence 710, App
43	17	2.0	307	9	US-09-878-178-710	Sequence 710, App
44	17	2.0	307	9	US-10-146-502-710	Sequence 710, App
45	17	2.0	307	9	US-10-066-543-3048	Sequence 3048, Ap
46	17	2.0	307	9	US-10-066-543-3351	Sequence 3351, Ap
47	17	2.0	357	9	US-09-345-373-79	Sequence 79, Appli
48	17	2.0	402	9	US-09-345-373-69	Sequence 69, Appli
49	17	2.0	426	9	US-09-345-373-95	Sequence 95, Appli
50	17	2.0	426	9	US-09-345-373-111	Sequence 111, App
51	17	2.0	426	9	US-09-345-373-119	Sequence 119, App
52	17	2.0	426	9	US-09-345-373-123	Sequence 123, App
53	17	2.0	426	9	US-09-345-373-127	Sequence 127, App
54	17	2.0	426	9	US-09-345-373-131	Sequence 131, App
55	17	2.0	426	9	US-09-345-373-135	Sequence 135, App
56	17	2.0	426	9	US-09-345-373-139	Sequence 139, App
57	17	2.0	426	9	US-09-345-373-145	Sequence 145, App
58	17	2.0	426	10	US-09-853-666-16	Sequence 16, Appli
59	17	2.0	426	10	US-09-853-666-32	Sequence 32, Appli
60	17	2.0	444	9	US-09-345-373-67	Sequence 67, Appli
61	17	2.0	444	9	US-10-194-443-3	Sequence 3, Appli
62	17	2.0	447	9	US-10-194-443-46	Sequence 46, Appli
63	17	2.0	451	9	US-09-918-995-15992	Sequence 15992, A
64	17	2.0	456	9	US-10-194-443-43	Sequence 43, Appli
65	17	2.0	456	9	US-09-345-373-115	Sequence 115, App
66	17	2.0	525	9	US-09-345-373-42	Sequence 42, Appli
67	17	2.0	525	9	US-09-345-373-65	Sequence 65, Appli
68	17	2.0	525	9	US-09-345-373-83	Sequence 83, Appli
69	17	2.0	525	9	US-09-345-373-84	Sequence 84, Appli
70	17	2.0	525	9	US-10-194-443-9	Sequence 9, Appli
71	17	2.0	543	10	US-09-974-300-540	Sequence 540, App
72	17	2.0	555	9	US-09-345-373-29	Sequence 29, Appli
73	17	2.0	627	9	US-09-345-373-38	Sequence 38, Appli
74	17	2.0	627	9	US-10-194-443-7	Sequence 7, Appli
75	17	2.0	663	10	US-09-974-300-7665	Sequence 7665, Ap
76	17	2.0	762	12	US-10-001-870-2	Sequence 2, Appli
77	17	2.0	779	10	US-09-772-134B-69	Sequence 69, Appli
78	17	2.0	828	10	US-09-862-810-3	Sequence 3, Appli
79	17	2.0	840	9	US-10-314-639-59	Sequence 59, Appli
80	17	2.0	840	12	US-10-059-964-59	Sequence 59, Appli
81	17	2.0	867	12	US-10-314-639-31	Sequence 31, Appli
82	17	2.0	867	12	US-10-059-964-31	Sequence 31, Appli
83	17	2.0	1130	10	US-09-841-786-9	Sequence 9, Appli
84	17	2.0	1503	9	US-09-938-842A-962	Sequence 962, App
85	17	2.0	1607	9	US-10-062-051-1	Sequence 1, Appli
86	17	2.0	1607	9	US-10-062-920-1	Sequence 1, Appli
87	17	2.0	1607	12	US-10-062-624-1	Sequence 1, Appli
88	17	2.0	1672	9	US-09-764-904-132	Sequence 132, App
89	17	2.0	1672	9	US-10-091-548-132	Sequence 132, App
90	17	2.0	1672	9	US-10-074-095-1193	Sequence 1193, Ap
91	17	2.0	1672	10	US-09-764-860-1193	Sequence 1193, Ap
92	17	2.0	1706	9	US-09-910-664-92	Sequence 92, Appli

c 93	17	2.0	2044	9	US-10-156-275-63	Sequence 63, Appl	c 166	16	1.9	1347	10	US-09-923-246-105	Sequence 105, App
c 94	17	2.0	2044	9	US-10-156-275-64	Sequence 64, Appl	c 167	16	1.9	1379	9	US-09-965-528-41	Sequence 41, Appl
c 95	17	2.0	2623	9	US-09-804-409A-7	Sequence 7, Appl	c 168	16	1.9	1463	9	US-10-106-989-1	Sequence 1, Appl
c 96	17	2.0	2733	9	US-10-128-714-214	Sequence 214, App	c 169	16	1.9	1463	10	US-09-350-874-1	Sequence 407, App
c 97	17	2.0	2733	9	US-10-128-714-5214	Sequence 5214, App	c 170	16	1.9	1489	9	US-09-764-868-407	Sequence 261, App
c 98	17	2.0	2780	10	US-09-841-786-10	Sequence 10, Appl	c 171	16	1.9	1489	9	US-10-103-313-261	Sequence 7, Appl
c 99	17	2.0	5895	7	US-08-781-986A-1	Sequence 1, Appl	c 172	16	1.9	1534	9	US-10-106-989-7	Sequence 7, Appl
c 100	17	2.0	9500	10	US-09-862-810-1	Sequence 1, Appl	c 173	16	1.9	1534	10	US-09-350-874-7	Sequence 9, Appl
c 101	17	2.0	9726	10	US-09-841-786-8	Sequence 8, Appl	c 174	16	1.9	1540	9	US-10-106-989-9	Sequence 9, Appl
c 102	17	2.0	9903	9	US-09-764-891-8188	Sequence 8188, Ap	c 175	16	1.9	1540	10	US-09-822-830A-413	Sequence 413, App
c 103	17	2.0	11330	10	US-09-841-786-15	Sequence 15, Appl	c 176	16	1.9	1564	10	US-10-092-154-1934	Sequence 1934, Ap
c 104	17	2.0	11474	9	US-10-092-154-1559	Sequence 1559, Ap	c 177	16	1.9	1617	9	US-09-764-847-1934	Sequence 1934, Ap
c 105	17	2.0	11474	10	US-09-764-847-1559	Sequence 1559, Ap	c 178	16	1.9	1617	10	US-09-764-847-1934	Sequence 1934, Ap
c 106	17	2.0	14117	9	US-09-954-531-1377	Sequence 1377, Ap	c 179	16	1.9	1617	10	US-09-764-877-3863	Sequence 3863, Ap
c 107	17	2.0	17450	9	US-09-764-891-8641	Sequence 8641, Ap	c 180	16	1.9	1726	9	US-10-007-270-10	Sequence 10, Appl
c 108	17	2.0	23419	9	US-10-091-504-1364	Sequence 1364, Ap	c 181	16	1.9	1781	10	US-09-818-512-1	Sequence 1, Appl
c 109	17	2.0	197997	10	US-09-822-246-3	Sequence 3, Appl	c 182	16	1.9	1831	9	US-10-037-270-939	Sequence 939, App
c 110	17	2.0	335913	9	US-09-754-853A-2	Sequence 2, Appl	c 183	16	1.9	1856	9	US-10-106-989-3	Sequence 3, Appl
c 111	17	2.0	335913	9	US-09-754-853A-3	Sequence 3, Appl	c 184	16	1.9	1856	10	US-09-350-874-3	Sequence 3, Appl
c 112	17	2.0	402850	9	US-09-844-653-5	Sequence 5, Appl	c 185	16	1.9	1907	10	US-10-106-989-5	Sequence 5, Appl
c 113	17	2.0	460681	10	US-09-933-2678A-1	Sequence 1, Appl	c 186	16	1.9	1907	10	US-09-350-874-5	Sequence 5, Appl
c 114	17	2.0	460681	10	US-09-790-988-1	Sequence 1, Appl	c 187	16	1.9	1916	9	US-09-822-846-179	Sequence 179, App
c 115	17	2.0	1691139	9	US-10-067-514-1	Sequence 1, Appl	c 188	16	1.9	1919	10	US-09-864-761-10504	Sequence 10504, A
c 116	17	2.0	1830121	9	US-10-329-960-1	Sequence 1, Appl	c 189	16	1.9	1923	9	US-09-938-842A-847	Sequence 847, App
c 117	17	2.0	3309400	9	US-09-738-626-1	Sequence 1, Appl	c 190	16	1.9	1931	9	US-10-114-893-173	Sequence 173, App
c 118	17	2.0	3309400	9	US-10-284-986-31	Sequence 31, Appl	c 191	16	1.9	1962	9	US-10-020-038-4	Sequence 4, Appl
c 119	16	1.9	20	10	US-09-846-808-31	Sequence 31, Appl	c 192	16	1.9	1976	9	US-10-158-846-9850	Sequence 9850, Ap
c 120	16	1.9	32	8	US-08-834-666A-56	Sequence 56, Appl	c 193	16	1.9	2035	9	US-09-764-891-8737	Sequence 8737, Ap
c 121	16	1.9	32	8	US-08-834-666A-61	Sequence 61, Appl	c 194	16	1.9	2061	9	US-09-842-758-23	Sequence 23, Appl
c 122	16	1.9	153	10	US-09-770-696-468	Sequence 468, App	c 195	16	1.9	2167	9	US-09-842-758-25	Sequence 25, Appl
c 123	16	1.9	208	10	US-09-864-761-30882	Sequence 30882, A	c 196	16	1.9	2355	9	US-09-842-758-21	Sequence 21, Appl
c 124	16	1.9	247	10	US-09-876-889-50	Sequence 50, Appl	c 197	16	1.9	2525	7	US-08-210-143-1	Sequence 1, Appl
c 125	16	1.9	270	10	US-09-878-574-14577	Sequence 14577, A	c 198	16	1.9	2699	8	US-08-834-666A-3	Sequence 3, Appl
c 126	16	1.9	270	10	US-09-878-574-15620	Sequence 15620, A	c 199	16	1.9	2729	10	US-09-822-849A-529	Sequence 529, App
c 127	16	1.9	313	10	US-09-964-824A-354	Sequence 354, App	c 200	16	1.9	2915	8	US-08-834-666A-5	Sequence 5, Appl
c 128	16	1.9	338	10	US-09-803-719-625	Sequence 625, App	c 201	16	1.9	3027	9	US-10-078-531-1	Sequence 1, Appl
c 129	16	1.9	362	9	US-10-060-036-958	Sequence 958, App	c 202	16	1.9	3115	9	US-09-877-650-1	Sequence 1, Appl
c 130	16	1.9	365	9	US-09-918-995-29967	Sequence 958, App	c 203	16	1.9	3115	10	US-09-871-856-1	Sequence 1, Appl
c 131	16	1.9	378	10	US-09-783-590-1730	Sequence 29967, A	c 204	16	1.9	3136	9	US-09-877-650-5	Sequence 5, Appl
c 132	16	1.9	382	9	US-09-991-936-514	Sequence 514, App	c 205	16	1.9	3136	10	US-09-871-856-5	Sequence 5, Appl
c 133	16	1.9	405	9	US-09-918-995-5190	Sequence 5190, App	c 206	16	1.9	3620	10	US-09-971-309-55	Sequence 55, Appl
c 134	16	1.9	416	9	US-09-918-995-12927	Sequence 12927, A	c 207	16	1.9	3668	9	US-10-007-270-8	Sequence 8, Appl
c 135	16	1.9	422	10	US-09-960-352-3153	Sequence 9153, App	c 208	16	1.9	3776	10	US-09-900-237-3	Sequence 3, Appl
c 136	16	1.9	431	10	US-10-106-698-888	Sequence 9337, App	c 209	16	1.9	3943	9	US-10-125-540-131	Sequence 131, App
c 137	16	1.9	431	10	US-09-878-574-5294	Sequence 888, App	c 210	16	1.9	3943	10	US-09-764-870-131	Sequence 131, App
c 138	16	1.9	433	10	US-09-960-352-10390	Sequence 5294, App	c 211	16	1.9	3945	9	US-10-125-540-242	Sequence 242, App
c 139	16	1.9	447	10	US-09-998-598-1736	Sequence 10390, A	c 212	16	1.9	3945	10	US-09-764-870-242	Sequence 242, App
c 140	16	1.9	456	10	US-09-998-598-1736	Sequence 1736, App	c 213	16	1.9	4334	9	US-10-187-319-38	Sequence 38, Appl
c 141	16	1.9	466	9	US-09-991-936-1335	Sequence 1335, App	c 214	16	1.9	4629	10	US-09-150-811-7	GENERAL INFORMA
c 142	16	1.9	483	10	US-09-864-761-14319	Sequence 14319, A	c 215	16	1.9	4999	9	US-10-007-968-14	Sequence 14, Appl
c 143	16	1.9	485	9	US-09-918-995-10008	Sequence 10008, A	c 216	16	1.9	4999	9	US-10-231-400-14	Sequence 14, Appl
c 144	16	1.9	487	9	US-09-918-995-24398	Sequence 24398, A	c 217	16	1.9	4999	10	US-09-740-211-14	Sequence 14, Appl
c 145	16	1.9	499	9	US-09-796-692-4216	Sequence 5387, App	c 218	16	1.9	5687	9	US-10-239-676-21	Sequence 21, Appl
c 146	16	1.9	499	9	US-10-040-862-5387	Sequence 5387, App	c 219	16	1.9	6402	9	US-10-187-319-36	Sequence 36, Appl
c 147	16	1.9	500	9	US-10-060-830-34	Sequence 34, Appl	c 220	16	1.9	7914	12	US-10-095-718-3	Sequence 3, Appl
c 148	16	1.9	503	9	US-09-918-995-22122	Sequence 22122, A	c 221	16	1.9	7944	12	US-10-095-718-1	Sequence 1, Appl
c 149	16	1.9	508	10	US-09-783-590-7777	Sequence 7777, App	c 222	16	1.9	8801	9	US-10-239-676-143	Sequence 143, App
c 150	16	1.9	587	10	US-09-864-761-7125	Sequence 7125, App	c 223	16	1.9	9009	9	US-09-957-641-1	Sequence 1, Appl
c 151	16	1.9	599	9	US-09-796-692-4216	Sequence 4216, App	c 224	16	1.9	9029	9	US-10-132-829-1	Sequence 1, Appl
c 152	16	1.9	603	9	US-09-910-664-47	Sequence 47, Appl	c 225	16	1.9	10758	7	US-08-781-986A-221	Sequence 221, App
c 153	16	1.9	643	9	US-09-899-495-37	Sequence 37, Appl	c 226	16	1.9	10953	7	US-08-781-986A-62	Sequence 62, Appl
c 154	16	1.9	705	10	US-09-995-598-77	Sequence 77, Appl	c 227	16	1.9	11933	9	US-10-007-968-13	Sequence 13, Appl
c 155	16	1.9	737	10	US-09-925-297-120	Sequence 120, App	c 228	16	1.9	11933	9	US-10-293-400-13	Sequence 13, Appl
c 156	16	1.9	741	9	US-09-809-391-202	Sequence 202, App	c 229	16	1.9	11933	10	US-09-740-211-13	Sequence 13, Appl
c 157	16	1.9	742	9	US-09-029-047-3	Sequence 3, Appl	c 230	16	1.9	15129	9	US-09-764-891-5895	Sequence 5895, App
c 158	16	1.9	773	9	US-09-809-391-20	Sequence 20, Appl	c 231	16	1.9	15299	9	US-09-736-457-1804	Sequence 1804, Ap
c 159	16	1.9	794	7	US-08-781-986A-835	Sequence 835, App	c 232	16	1.9	17569	9	US-09-902-941-1804	Sequence 1804, Ap
c 160	16	1.9	900	9	US-10-302-557-1	Sequence 1, Appl	c 233	16	1.9	17569	9	US-09-849-626-1804	Sequence 1804, Ap
c 161	16	1.9	981	10	US-09-881-752A-283	Sequence 283, App	c 234	16	1.9	17569	9	US-10-017-754-1804	Sequence 1804, Ap
c 162	16	1.9	1064	10	US-09-764-877-3862	Sequence 3862, App	c 235	16	1.9	53000	9	US-09-953-611-10	Sequence 10, Appl
c 163	16	1.9	1141	7	US-08-781-986A-526	Sequence 526, App	c 236	16	1.9	116592	10	US-09-818-512-3	Sequence 3, Appl
c 164	16	1.9	1141	7	US-08-781-986A-526	Sequence 526, App	c 237	16	1.9	116592	10	US-09-818-512-3	Sequence 3, Appl
c 165	16	1.9	1141	7	US-08-781-986A-526	Sequence 526, App	c 238	16	1.9	126512	10	US-09-804-474A-3	Sequence 3, Appl

239	16	1.9	132762	9	US-09-954-556-17	Sequence 17, Appl	312	15	1.8	387	10	US-09-878-574-1081	Sequence 1081, Ap
240	16	1.9	186957	9	US-10-185-770-3	Sequence 3, Appl	c 313	15	1.8	390	9	US-09-736-457-1685	Sequence 1685, Ap
c 241	16	1.9	197997	10	US-09-822-246-3	Sequence 3, Appl	c 314	15	1.8	390	9	US-09-902-941-1685	Sequence 1685, Ap
c 242	16	1.9	202001	9	US-10-274-990-3	Sequence 3, Appl	c 315	15	1.8	390	9	US-09-849-626-1685	Sequence 1685, Ap
c 243	16	1.9	202001	10	US-09-734-674-3	Sequence 3, Appl	c 316	15	1.8	390	9	US-10-017-754-1685	Sequence 1685, Ap
c 244	16	1.9	397658	10	US-09-813-320-3	Sequence 3, Appl	c 317	15	1.8	401	9	US-09-946-807-1085	Sequence 1085, Ap
c 245	16	1.9	402850	9	US-09-844-653-5	Sequence 5, Appl	c 318	15	1.8	401	9	US-09-946-807-1096	Sequence 1096, Ap
c 246	16	1.9	465237	10	US-09-933-267A-1	Sequence 1, Appl	c 319	15	1.8	401	10	US-09-795-668-1085	Sequence 1085, Ap
c 247	16	1.9	513509	9	US-09-754-853A-4	Sequence 1, Appl	c 320	15	1.8	401	10	US-09-795-668-1096	Sequence 1096, Ap
c 248	16	1.9	1503841	9	US-09-946-807-1	Sequence 1, Appl	c 321	15	1.8	401	10	US-09-795-668-1096	Sequence 1096, Ap
c 249	16	1.9	1503841	10	US-09-946-807-1	Sequence 1, Appl	c 322	15	1.8	401	9	US-09-795-668-1096	Sequence 1096, Ap
c 250	16	1.9	1503841	10	US-09-795-668-1	Sequence 1, Appl	c 323	15	1.8	405	9	US-09-946-807-1453	Sequence 1453, Ap
c 251	16	1.9	1503841	10	US-09-795-668-1	Sequence 1, Appl	c 324	15	1.8	405	10	US-09-795-668-1453	Sequence 1453, Ap
c 252	16	1.9	1503841	10	US-09-795-668-1	Sequence 1, Appl	c 325	15	1.8	405	10	US-09-795-668-1453	Sequence 1453, Ap
c 253	16	1.9	1503841	10	US-09-795-668-1	Sequence 1, Appl	c 326	15	1.8	406	9	US-09-918-995-17263	Sequence 17263, A
c 254	15	1.8	24	9	US-09-927-161-14	Sequence 14, Appl	c 327	15	1.8	409	9	US-09-803-719-361	Sequence 361, App
c 255	15	1.8	25	9	US-10-098-263B-85535	Sequence 85535, A	c 328	15	1.8	416	9	US-09-918-995-17755	Sequence 17755, A
c 256	15	1.8	37	9	US-10-156-306-2676	Sequence 2676, Ap	c 329	15	1.8	417	9	US-09-918-995-16856	Sequence 16856, A
c 257	15	1.8	47	9	US-09-345-373-51	Sequence 51, Appl	c 330	15	1.8	424	9	US-09-960-352-981	Sequence 981, App
c 258	15	1.8	47	9	US-10-194-443-31	Sequence 31, Appl	c 331	15	1.8	424	9	US-09-918-995-4522	Sequence 4522, Ap
c 259	15	1.8	48	9	US-09-345-373-47	Sequence 47, Appl	c 332	15	1.8	424	10	US-09-871-916-3	Sequence 3, Appl
c 260	15	1.8	48	9	US-10-194-443-37	Sequence 27, Appl	c 333	15	1.8	424	10	US-09-960-352-280	Sequence 280, App
c 261	15	1.8	117	10	US-09-864-761-30736	Sequence 30736, A	c 334	15	1.8	430	10	US-09-960-352-3824	Sequence 3824, Ap
c 262	15	1.8	149	10	US-09-864-761-27904	Sequence 27904, A	c 335	15	1.8	433	10	US-09-962-436-541	Sequence 541, App
c 263	15	1.8	151	9	US-10-322-149-7	Sequence 7, Appl	c 336	15	1.8	433	10	US-09-964-824A-234	Sequence 234, App
c 264	15	1.8	201	10	US-09-864-761-23168	Sequence 23168, A	c 337	15	1.8	434	10	US-09-920-300A-1734	Sequence 1734, Ap
c 265	15	1.8	203	9	US-09-796-692-7992	Sequence 7992, Ap	c 338	15	1.8	434	12	US-10-033-528-1734	Sequence 1734, Ap
c 266	15	1.8	203	9	US-10-040-862-7992	Sequence 7992, Ap	c 339	15	1.8	435	10	US-09-954-456-2128	Sequence 2128, Ap
c 267	15	1.8	223	9	US-10-015-219-1251	Sequence 1251, Ap	c 340	15	1.8	435	10	US-09-880-107-292	Sequence 292, App
c 268	15	1.8	223	10	US-09-777-564-1251	Sequence 1251, Ap	c 341	15	1.8	436	9	US-09-918-995-37143	Sequence 37143, A
c 269	15	1.8	231	10	US-09-974-300-7969	Sequence 7969, Ap	c 342	15	1.8	441	10	US-09-960-352-12356	Sequence 12356, A
c 270	15	1.8	231	10	US-09-974-300-82235	Sequence 8235, Ap	c 343	15	1.8	442	10	US-09-867-701-4545	Sequence 4545, Ap
c 271	15	1.8	234	10	US-09-911-935A-28	Sequence 28, Appl	c 344	15	1.8	455	9	US-09-918-995-2683	Sequence 2683, Ap
c 272	15	1.8	257	10	US-09-878-574-11857	Sequence 11857, A	c 345	15	1.8	456	10	US-09-960-352-759	Sequence 759, App
c 273	15	1.8	261	10	US-09-878-574-11251	Sequence 11251, A	c 346	15	1.8	457	9	US-09-918-995-12862	Sequence 12862, A
c 274	15	1.8	262	9	US-09-918-995-30578	Sequence 30578, A	c 347	15	1.8	457	9	US-09-918-995-26848	Sequence 26848, A
c 275	15	1.8	263	10	US-09-964-824A-137	Sequence 137, App	c 348	15	1.8	458	9	US-09-918-995-12779	Sequence 12779, A
c 276	15	1.8	269	10	US-09-969-373-1341	Sequence 1341, Ap	c 349	15	1.8	462	9	US-09-918-995-13048	Sequence 13048, A
c 277	15	1.8	271	10	US-09-878-574-8314	Sequence 8314, Ap	c 350	15	1.8	464	9	US-09-918-995-9991	Sequence 9991, Ap
c 278	15	1.8	272	10	US-09-878-574-8750	Sequence 8750, Ap	c 351	15	1.8	465	9	US-09-918-995-35399	Sequence 35399, A
c 279	15	1.8	273	10	US-09-815-242-1793	Sequence 1793, Ap	c 352	15	1.8	466	9	US-09-918-995-21309	Sequence 21309, A
c 280	15	1.8	276	9	US-10-083-357-32	Sequence 32, Appl	c 353	15	1.8	466	10	US-09-864-761-6578	Sequence 6578, Ap
c 281	15	1.8	278	9	US-10-102-524-184	Sequence 184, App	c 354	15	1.8	467	10	US-09-563-817-719	Sequence 719, App
c 282	15	1.8	284	10	US-09-783-590-6660	Sequence 6660, Ap	c 355	15	1.8	471	10	US-09-983-965-5525	Sequence 5525, Ap
c 283	15	1.8	285	10	US-09-864-761-28013	Sequence 28013, A	c 356	15	1.8	473	9	US-09-918-995-17063	Sequence 17063, A
c 284	15	1.8	289	10	US-09-960-352-5943	Sequence 5943, Ap	c 357	15	1.8	474	10	US-09-864-761-1891	Sequence 1891, Ap
c 285	15	1.8	298	10	US-09-878-574-14594	Sequence 14594, A	c 358	15	1.8	474	10	US-09-864-761-14059	Sequence 14059, A
c 286	15	1.8	299	10	US-09-879-536-349	Sequence 349, App	c 359	15	1.8	475	10	US-09-867-701-3078	Sequence 3078, Ap
c 287	15	1.8	307	10	US-09-728-446-932	Sequence 932, App	c 360	15	1.8	476	9	US-09-918-995-2041	Sequence 2041, Ap
c 288	15	1.8	313	9	US-10-060-036-1780	Sequence 1780, Ap	c 361	15	1.8	476	9	US-09-918-995-17912	Sequence 17912, A
c 289	15	1.8	317	9	US-09-803-719-1970	Sequence 1970, Ap	c 362	15	1.8	477	9	US-09-918-995-1894	Sequence 1894, Ap
c 290	15	1.8	330	10	US-09-294-0938-122	Sequence 122, App	c 363	15	1.8	477	9	US-09-918-995-16654	Sequence 16654, A
c 291	15	1.8	334	10	US-09-983-965-3857	Sequence 3857, App	c 364	15	1.8	477	9	US-09-918-995-21316	Sequence 21316, A
c 292	15	1.8	337	10	US-09-983-965-2463	Sequence 2463, Ap	c 365	15	1.8	479	10	US-09-864-761-4010	Sequence 4010, Ap
c 293	15	1.8	341	9	US-09-899-495-5	Sequence 5, Appl	c 366	15	1.8	481	9	US-09-918-995-6811	Sequence 6811, Ap
c 294	15	1.8	346	10	US-09-878-574-2152	Sequence 2152, Ap	c 367	15	1.8	482	9	US-09-918-995-16965	Sequence 16965, A
c 295	15	1.8	348	10	US-09-878-574-3842	Sequence 3842, Ap	c 368	15	1.8	482	10	US-09-920-300A-1363	Sequence 1363, Ap
c 296	15	1.8	351	9	US-09-920-455-20	Sequence 20, Appl	c 369	15	1.8	483	12	US-10-033-528-1363	Sequence 1363, Ap
c 297	15	1.8	360	10	US-09-924-401-56	Sequence 56, Appl	c 370	15	1.8	483	9	US-09-796-692-4807	Sequence 4807, Ap
c 298	15	1.8	363	10	US-09-783-590-10992	Sequence 10992, A	c 371	15	1.8	483	9	US-10-040-862-4807	Sequence 4807, Ap
c 299	15	1.8	364	10	US-09-878-574-182	Sequence 182, App	c 372	15	1.8	483	10	US-09-960-352-10904	Sequence 10904, A
c 300	15	1.8	365	9	US-09-918-995-17926	Sequence 17926, A	c 373	15	1.8	485	10	US-09-864-761-15175	Sequence 15175, A
c 301	15	1.8	367	9	US-10-106-698-2450	Sequence 2450, Ap	c 374	15	1.8	486	9	US-09-918-995-26813	Sequence 26813, A
c 302	15	1.8	368	10	US-09-864-761-11417	Sequence 11417, A	c 375	15	1.8	486	10	US-09-864-761-6453	Sequence 6453, Ap
c 303	15	1.8	368	10	US-09-864-761-14006	Sequence 14006, A	c 376	15	1.8	486	10	US-09-815-242-9265	Sequence 9265, Ap
c 304	15	1.8	369	10	US-09-954-456-1688	Sequence 1688, Ap	c 377	15	1.8	486	10	US-09-918-995-25184	Sequence 25184, A
c 305	15	1.8	375	9	US-09-918-995-29705	Sequence 29705, A	c 378	15	1.8	487	9	US-09-918-995-9023	Sequence 9023, Ap
c 306	15	1.8	378	10	US-09-867-701-986	Sequence 986, App	c 379	15	1.8	488	9	US-09-918-995-37876	Sequence 37876, A
c 307	15	1.8	380	10	US-09-745-288-31	Sequence 31, Appl	c 380	15	1.8	488	12	US-10-033-528-1793	Sequence 1793, Ap
c 308	15	1.8	382	9	US-09-918-995-16544	Sequence 16544, A	c 381	15	1.8	488	12	US-09-918-995-37876	Sequence 37876, A
c 309	15	1.8	382	10	US-09-770-791-161	Sequence 161, App	c 382	15	1.8	491	9	US-09-922-758-8	Sequence 8, Appl
c 310	15	1.8	385	9	US-09-920-455-273	Sequence 273, App	c 383	15	1.8	494	9	US-10-060-036-1001	Sequence 1001, Ap
c 311	15	1.8	386	9	US-10-102-524-1204	Sequence 1204, Ap	c 384	15	1.8	500	10	US-09-864-761-11395	Sequence 11395, A



C 531	15	1.8	755	10	US-09-989-731-15	Sequence 15, Appl	604	15	1.8	1333	10	US-09-757-982-13	Sequence 13, Appl
C 532	15	1.8	755	10	US-09-989-732-15	Sequence 15, Appl	605	15	1.8	1351	9	US-10-153-668-347	Sequence 347, App
C 533	15	1.8	755	10	US-09-991-073-15	Sequence 15, Appl	606	15	1.8	1358	10	US-09-939-293-1	Sequence 1, Appl
C 534	15	1.8	755	10	US-09-990-442-15	Sequence 15, Appl	C 607	15	1.8	1397	10	US-09-974-300-561	Sequence 561, App
C 535	15	1.8	755	10	US-09-991-163-15	Sequence 15, Appl	C 608	15	1.8	1404	9	US-10-037-270-710	Sequence 710, App
C 536	15	1.8	755	10	US-09-993-604-15	Sequence 15, Appl	C 609	15	1.8	1460	9	US-09-822-846-490	Sequence 490, App
C 537	15	1.8	755	10	US-09-990-456-15	Sequence 15, Appl	C 610	15	1.8	1486	9	US-10-037-270-193	Sequence 193, App
C 538	15	1.8	755	10	US-09-989-721-15	Sequence 15, Appl	C 611	15	1.8	1503	9	US-10-198-846-14076	Sequence 14076, A
C 539	15	1.8	768	9	US-09-989-720-707	Sequence 1307, App	C 612	15	1.8	1510	9	US-10-013-313-76	Sequence 76, Appl
C 540	15	1.8	792	9	US-09-938-842A-1320	Sequence 1320, Ap	C 613	15	1.8	1515	9	US-10-014-101-30	Sequence 30, Appl
C 541	15	1.8	794	9	US-10-043-487-53	Sequence 53, Appl	C 614	15	1.8	1530	9	US-10-198-846-10434	Sequence 10434, A
C 542	15	1.8	806	9	US-09-954-531-1381	Sequence 1381, Ap	C 615	15	1.8	1531	10	US-09-070-927A-412	Sequence 412, App
C 543	15	1.8	806	10	US-09-954-456-1833	Sequence 1833, Ap	C 616	15	1.8	1538	10	US-09-925-731-3	Sequence 3, Appl
C 544	15	1.8	825	9	US-10-198-846-3041	Sequence 3041, Ap	C 617	15	1.8	1566	9	US-10-081-051-59	Sequence 59, Appl
C 545	15	1.8	837	9	US-09-934-455-53	Sequence 53, Appl	C 618	15	1.8	1566	9	US-09-769-787-341	Sequence 341, App
C 546	15	1.8	838	10	US-09-770-445-668	Sequence 668, App	C 619	15	1.8	1581	10	US-09-934-868-59	Sequence 59, Appl
C 547	15	1.8	845	9	US-10-322-149-15	Sequence 15, Appl	C 620	15	1.8	1584	10	US-09-815-242-7277	Sequence 7277, Ap
C 548	15	1.8	875	7	US-08-781-986A-319	Sequence 319, App	C 621	15	1.8	1592	9	US-10-106-698-978	Sequence 978, App
C 549	15	1.8	911	10	US-09-925-297-132	Sequence 132, App	C 622	15	1.8	1605	9	US-09-938-842A-4867	Sequence 4867, Ap
C 550	15	1.8	916	10	US-09-770-445-432	Sequence 432, App	C 623	15	1.8	1619	10	US-09-953-956-1	Sequence 1, Appl
C 551	15	1.8	917	10	US-09-770-445-427	Sequence 427, App	C 624	15	1.8	1619	12	US-10-114-464-1	Sequence 1, Appl
C 552	15	1.8	919	9	US-09-938-842A-4314	Sequence 4314, Ap	C 625	15	1.8	1622	9	US-09-822-846-407	Sequence 407, App
C 553	15	1.8	921	9	US-10-106-698-174	Sequence 174, App	C 626	15	1.8	1622	10	US-09-925-297-176	Sequence 176, App
C 554	15	1.8	922	9	US-10-074-095-1112	Sequence 1112, Ap	C 627	15	1.8	1653	9	US-09-738-626-77	Sequence 77, Appl
C 555	15	1.8	922	10	US-09-764-860-1112	Sequence 1112, Ap	C 628	15	1.8	1669	9	US-09-822-846-300	Sequence 300, App
C 556	15	1.8	925	10	US-09-764-877-3278	Sequence 3277, Ap	C 629	15	1.8	1671	12	US-10-044-090-462	Sequence 462, App
C 557	15	1.8	925	10	US-09-764-877-3278	Sequence 3278, Ap	C 630	15	1.8	1726	12	US-10-062-254-327	Sequence 327, App
C 558	15	1.8	926	10	US-09-939-980-118	Sequence 118, App	C 631	15	1.8	1740	10	US-09-938-842A-2315	Sequence 2315, Ap
C 559	15	1.8	936	9	US-10-198-846-4144	Sequence 4144, Ap	C 632	15	1.8	1756	10	US-09-925-300-753	Sequence 753, App
C 560	15	1.8	942	10	US-09-974-300-5269	Sequence 5269, Ap	C 633	15	1.8	1784	9	US-09-878-131-1	Sequence 1, Appl
C 561	15	1.8	958	9	US-10-126-139-5	Sequence 5, Appl	C 634	15	1.8	1784	9	US-09-878-131-3	Sequence 3, Appl
C 562	15	1.8	958	9	US-10-126-798-5	Sequence 5, Appl	C 635	15	1.8	1821	10	US-09-894-633A-100	Sequence 100, App
C 563	15	1.8	958	9	US-09-808-898-5	Sequence 5, Appl	C 636	15	1.8	1839	10	US-09-815-242-7198	Sequence 7198, Ap
C 564	15	1.8	958	9	US-10-126-777-5	Sequence 5, Appl	C 637	15	1.8	1916	10	US-09-925-297-296	Sequence 296, App
C 565	15	1.8	958	10	US-09-803-211-5	Sequence 5, Appl	C 638	15	1.8	1931	9	US-10-062-848-4	Sequence 4, Appl
C 566	15	1.8	958	10	US-09-746-485A-5	Sequence 5, Appl	C 639	15	1.8	1936	9	US-10-014-101-11	Sequence 11, Appl
C 567	15	1.8	960	9	US-10-106-698-45	Sequence 45, Appl	C 640	15	1.8	1939	9	US-09-895-913A-261	Sequence 261, App
C 568	15	1.8	965	9	US-09-883-152-9	Sequence 9, Appl	C 641	15	1.8	1939	9	US-09-883-152-22	Sequence 22, Appl
C 569	15	1.8	966	10	US-09-864-761-2514	Sequence 2514, Ap	C 642	15	1.8	1957	9	US-09-892-877-91	Sequence 91, Appl
C 570	15	1.8	968	7	US-08-781-986A-563	Sequence 563, App	C 643	15	1.8	1957	9	US-09-948-783-90	Sequence 90, Appl
C 571	15	1.8	987	9	US-10-198-846-14073	Sequence 14073, A	C 644	15	1.8	1965	9	US-09-938-842A-1948	Sequence 1948, Ap
C 572	15	1.8	1035	9	US-09-938-842A-1734	Sequence 1734, Ap	C 645	15	1.8	1971	10	US-09-864-761-10277	Sequence 10277, A
C 573	15	1.8	1059	10	US-09-815-242-4268	Sequence 4268, Ap	C 646	15	1.8	2000	9	US-09-938-842A-2946	Sequence 2946, Ap
C 574	15	1.8	1077	10	US-09-770-445-129	Sequence 129, App	C 647	15	1.8	2000	9	US-09-938-842A-3195	Sequence 3195, Ap
C 575	15	1.8	1090	7	US-08-781-986A-9	Sequence 9, Appl	C 648	15	1.8	2000	9	US-09-938-842A-3293	Sequence 3293, Ap
C 576	15	1.8	1092	10	US-09-815-242-8481	Sequence 8481, Ap	C 649	15	1.8	2000	9	US-09-938-842A-3574	Sequence 3574, Ap
C 577	15	1.8	1100	10	US-09-263-959-265	Sequence 265, App	C 650	15	1.8	2000	9	US-09-938-842A-3606	Sequence 3606, Ap
C 578	15	1.8	1104	9	US-10-286-264-99	Sequence 265, App	C 651	15	1.8	2000	9	US-09-938-842A-3617	Sequence 3617, Ap
C 579	15	1.8	1108	10	US-09-070-927A-523	Sequence 523, App	C 652	15	1.8	2000	9	US-09-938-842A-3686	Sequence 3686, Ap
C 580	15	1.8	1110	9	US-09-883-152-7	Sequence 7, Appl	C 653	15	1.8	2000	9	US-09-938-842A-3877	Sequence 3877, Ap
C 581	15	1.8	1117	10	US-09-881-752A-221	Sequence 221, App	C 654	15	1.8	2000	9	US-09-938-842A-3974	Sequence 3974, Ap
C 582	15	1.8	1178	10	US-09-840-787-88	Sequence 88, Appl	C 655	15	1.8	2000	9	US-09-938-842A-3990	Sequence 3990, Ap
C 583	15	1.8	1194	10	US-09-815-242-6753	Sequence 6753, Ap	C 656	15	1.8	2000	9	US-09-938-842A-4195	Sequence 4195, Ap
C 584	15	1.8	1195	9	US-10-198-846-10962	Sequence 10962, A	C 657	15	1.8	2000	9	US-09-938-842A-4222	Sequence 4222, Ap
C 585	15	1.8	1195	9	US-10-205-823-259	Sequence 259, App	C 658	15	1.8	2000	9	US-09-938-842A-4305	Sequence 4305, Ap
C 586	15	1.8	1200	10	US-09-887-576-761	Sequence 761, App	C 659	15	1.8	2000	9	US-09-938-842A-4537	Sequence 4537, Ap
C 587	15	1.8	1200	10	US-09-925-300-488	Sequence 488, App	C 660	15	1.8	2000	9	US-09-938-842A-4817	Sequence 4817, Ap
C 588	15	1.8	1208	12	US-10-044-090-84	Sequence 84, Appl	C 661	15	1.8	2000	10	US-09-887-576-93	Sequence 93, Appl
C 589	15	1.8	1209	10	US-09-815-242-7437	Sequence 7437, Ap	C 662	15	1.8	2000	10	US-09-887-576-151	Sequence 151, App
C 590	15	1.8	1227	10	US-09-822-830A-18	Sequence 18, Appl	C 663	15	1.8	2004	10	US-09-887-576-317	Sequence 317, App
C 591	15	1.8	1230	10	US-09-815-242-4310	Sequence 4310, Ap	C 664	15	1.8	2007	10	US-09-815-242-3963	Sequence 3963, App
C 592	15	1.8	1245	9	US-09-938-842A-2546	Sequence 2546, Ap	C 665	15	1.8	2017	9	US-10-198-846-10186	Sequence 10186, A
C 593	15	1.8	1245	10	US-09-815-242-8431	Sequence 8431, Ap	C 666	15	1.8	2024	9	US-09-374-046A-111	Sequence 111, App
C 594	15	1.8	1245	10	US-09-815-242-8712	Sequence 8712, Ap	C 667	15	1.8	2037	10	US-09-815-242-3891	Sequence 3891, Ap
C 595	15	1.8	1245	10	US-09-956-398A-1	Sequence 1, Appl	C 668	15	1.8	2055	9	US-10-081-119-25	Sequence 25, Appl
C 596	15	1.8	1250	9	US-09-466-734A-15	Sequence 15, Appl	C 669	15	1.8	2055	10	US-09-804-060-1	Sequence 1, Appl
C 597	15	1.8	1251	9	US-10-198-846-13243	Sequence 13243, A	C 670	15	1.8	2078	9	US-10-091-572-812	Sequence 812, App
C 598	15	1.8	1275	9	US-09-938-842A-5037	Sequence 5037, A	C 671	15	1.8	2078	9	US-09-764-891-9236	Sequence 9236, Ap
C 599	15	1.8	1284	10	US-09-815-242-9345	Sequence 9345, Ap	C 672	15	1.8	2082	9	US-09-822-846-515	Sequence 515, App
C 600	15	1.8	1284	10	US-09-815-242-9542	Sequence 9542, Ap	C 673	15	1.8	2139	9	US-09-938-842A-2652	Sequence 2652, Ap
C 601	15	1.8	1288	9	US-09-938-842A-4207	Sequence 4207, Ap	C 674	15	1.8	2141	10	US-09-841-786-11	Sequence 11, Appl
C 602	15	1.8	1295	9	US-09-510-332-100	Sequence 100, App	C 675	15	1.8	2148	10	US-09-925-301-247	Sequence 247, App
C 603	15	1.8	1332	10	US-09-757-982-15	Sequence 15, Appl	C 676	15	1.8	2160	9	US-09-925-299-160	Sequence 160, App

C 677	15	1.8	2160	9	US-10-106-698-303	Sequence 303, App	C 750	15	1.8	2886	9	US-10-176-492-7	Sequence 7, Appli
C 678	15	1.8	2160	10	US-09-925-299-160	Sequence 160, App	C 751	15	1.8	2886	9	US-10-176-747-7	Sequence 7, Appli
C 679	15	1.8	2179	9	US-10-138-075-3	Sequence 3, Appli	C 752	15	1.8	2886	9	US-10-176-750-7	Sequence 7, Appli
C 680	15	1.8	2199	9	US-10-277-156-3	Sequence 3, Appli	C 753	15	1.8	2886	9	US-10-176-985-7	Sequence 7, Appli
C 681	15	1.8	2231	9	US-10-106-698-1260	Sequence 1260, Ap	C 754	15	1.8	2886	9	US-10-176-987-7	Sequence 7, Appli
C 682	15	1.8	2250	10	US-09-925-301-209	Sequence 209, App	C 755	15	1.8	2886	9	US-10-176-991-7	Sequence 7, Appli
C 683	15	1.8	2257	10	US-09-972-912-42	Sequence 42, Appli	C 756	15	1.8	2886	9	US-10-176-992-7	Sequence 7, Appli
C 684	15	1.8	2268	10	US-09-887-576-669	Sequence 669, App	C 757	15	1.8	2886	9	US-10-176-993-7	Sequence 7, Appli
C 685	15	1.8	2302	10	US-09-729-674-161	Sequence 161, App	C 758	15	1.8	2886	9	US-10-184-658-7	Sequence 7, Appli
C 686	15	1.8	2323	7	US-10-024-632-3	Sequence 3, Appli	C 759	15	1.8	2886	9	US-09-990-711-13	Sequence 13, Appli
C 687	15	1.8	2332	9	US-08-781-986A-477	Sequence 477, App	C 760	15	1.8	2886	9	US-10-173-695-7	Sequence 7, Appli
C 688	15	1.8	2338	12	US-10-044-090-69	Sequence 69, Appli	C 761	15	1.8	2886	9	US-10-173-697-7	Sequence 7, Appli
C 689	15	1.8	2391	10	US-09-764-877-3666	Sequence 3666, Ap	C 762	15	1.8	2886	9	US-10-173-705-7	Sequence 7, Appli
C 690	15	1.8	2401	9	US-10-036-492-14	Sequence 14, Appli	C 763	15	1.8	2886	9	US-10-174-576-7	Sequence 7, Appli
C 691	15	1.8	2406	9	US-10-157-031-368	Sequence 368, App	C 764	15	1.8	2886	9	US-10-174-585-7	Sequence 7, Appli
C 692	15	1.8	2413	10	US-09-917-800A-1688	Sequence 1688, Ap	C 765	15	1.8	2886	9	US-10-174-586-7	Sequence 7, Appli
C 693	15	1.8	2426	7	US-08-781-986A-542	Sequence 542, App	C 766	15	1.8	2886	9	US-10-175-747-7	Sequence 7, Appli
C 694	15	1.8	2434	9	US-10-036-492-9	Sequence 9, Appli	C 767	15	1.8	2886	9	US-10-176-481-7	Sequence 7, Appli
C 695	15	1.8	2481	9	US-10-027-049-1	Sequence 1, Appli	C 768	15	1.8	2886	9	US-10-176-485-7	Sequence 7, Appli
C 696	15	1.8	2487	9	US-10-270-336-1	Sequence 27, Appli	C 769	15	1.8	2886	9	US-10-176-487-7	Sequence 7, Appli
C 697	15	1.8	2512	9	US-10-036-492-27	Sequence 27, Appli	C 770	15	1.8	2886	9	US-10-176-493-7	Sequence 7, Appli
C 698	15	1.8	2538	9	US-10-152-661-511	Sequence 511, App	C 771	15	1.8	2886	9	US-10-176-756-7	Sequence 7, Appli
C 699	15	1.8	2538	9	US-09-866-050A-511	Sequence 511, App	C 772	15	1.8	2886	9	US-10-176-911-7	Sequence 7, Appli
C 700	15	1.8	2565	10	US-09-801-368-61	Sequence 61, Appli	C 773	15	1.8	2886	9	US-10-176-919-7	Sequence 7, Appli
C 701	15	1.8	2574	9	US-09-938-842A-673	Sequence 673, Appli	C 774	15	1.8	2886	9	US-10-176-925-7	Sequence 7, Appli
C 702	15	1.8	2583	10	US-09-815-242-6726	Sequence 6726, Ap	C 775	15	1.8	2886	9	US-10-176-978-7	Sequence 7, Appli
C 703	15	1.8	2593	9	US-10-001-857-67	Sequence 67, Appli	C 776	15	1.8	2886	9	US-10-179-510-7	Sequence 7, Appli
C 704	15	1.8	2604	9	US-10-027-049-3	Sequence 3, Appli	C 777	15	1.8	2886	9	US-10-180-543-7	Sequence 7, Appli
C 705	15	1.8	2619	10	US-09-815-242-7466	Sequence 7466, Ap	C 778	15	1.8	2886	9	US-10-180-544-7	Sequence 7, Appli
C 706	15	1.8	2633	10	US-09-764-877-2154	Sequence 2154, Ap	C 779	15	1.8	2886	9	US-10-180-546-7	Sequence 7, Appli
C 707	15	1.8	2655	9	US-10-153-668-287	Sequence 287, App	C 780	15	1.8	2886	9	US-10-180-547-7	Sequence 7, Appli
C 708	15	1.8	2674	10	US-09-764-877-3180	Sequence 3180, Ap	C 781	15	1.8	2886	9	US-10-180-549-7	Sequence 7, Appli
C 709	15	1.8	2760	10	US-09-922-261-359	Sequence 359, App	C 782	15	1.8	2886	9	US-10-180-555-7	Sequence 7, Appli
C 710	15	1.8	2769	9	US-10-328-459-1	Sequence 1, Appli	C 783	15	1.8	2886	9	US-10-180-559-7	Sequence 7, Appli
C 711	15	1.8	2776	9	US-10-091-504-1907	Sequence 1907, Ap	C 784	15	1.8	2886	9	US-10-181-000-7	Sequence 7, Appli
C 712	15	1.8	2776	9	US-09-764-891-6805	Sequence 6805, Ap	C 785	15	1.8	2886	9	US-10-183-010-7	Sequence 7, Appli
C 713	15	1.8	2776	10	US-09-764-869-1907	Sequence 1907, Ap	C 786	15	1.8	2886	9	US-10-183-012-7	Sequence 7, Appli
C 714	15	1.8	2879	9	US-10-153-668-147	Sequence 147, App	C 787	15	1.8	2886	9	US-10-184-614-7	Sequence 7, Appli
C 715	15	1.8	2886	9	US-09-992-598-13	Sequence 13, Appli	C 788	15	1.8	2886	9	US-10-184-623-7	Sequence 7, Appli
C 716	15	1.8	2886	9	US-09-989-293A-13	Sequence 13, Appli	C 789	15	1.8	2886	9	US-10-184-635-7	Sequence 7, Appli
C 717	15	1.8	2886	9	US-09-989-745-13	Sequence 13, Appli	C 790	15	1.8	2886	9	US-10-184-637-7	Sequence 7, Appli
C 718	15	1.8	2886	9	US-09-990-444-13	Sequence 13, Appli	C 791	15	1.8	2886	9	US-10-184-646-7	Sequence 7, Appli
C 719	15	1.8	2886	9	US-09-989-730-13	Sequence 13, Appli	C 792	15	1.8	2886	9	US-10-184-647-7	Sequence 7, Appli
C 720	15	1.8	2886	9	US-09-990-436-13	Sequence 13, Appli	C 793	15	1.8	2886	9	US-10-184-652-7	Sequence 7, Appli
C 721	15	1.8	2886	9	US-09-991-181-13	Sequence 13, Appli	C 794	15	1.8	2886	9	US-10-187-584-7	Sequence 7, Appli
C 722	15	1.8	2886	9	US-09-993-687-13	Sequence 13, Appli	C 795	15	1.8	2886	9	US-10-187-596-7	Sequence 7, Appli
C 723	15	1.8	2886	9	US-09-989-734-13	Sequence 13, Appli	C 796	15	1.8	2886	9	US-10-187-745-7	Sequence 7, Appli
C 724	15	1.8	2886	9	US-09-997-653-13	Sequence 13, Appli	C 797	15	1.8	2886	9	US-10-187-885-7	Sequence 7, Appli
C 725	15	1.8	2886	9	US-10-174-590-7	Sequence 7, Appli	C 798	15	1.8	2886	9	US-10-187-886-7	Sequence 7, Appli
C 726	15	1.8	2886	9	US-10-176-758-7	Sequence 7, Appli	C 799	15	1.8	2886	9	US-10-199-464-7	Sequence 7, Appli
C 727	15	1.8	2886	9	US-10-173-737-7	Sequence 7, Appli	C 800	15	1.8	2886	9	US-10-176-751-7	Sequence 7, Appli
C 728	15	1.8	2886	9	US-09-993-667-13	Sequence 13, Appli	C 801	15	1.8	2886	9	US-10-176-760-7	Sequence 7, Appli
C 729	15	1.8	2886	9	US-10-173-706-7	Sequence 7, Appli	C 802	15	1.8	2886	9	US-10-176-990-7	Sequence 7, Appli
C 730	15	1.8	2886	9	US-10-175-738-7	Sequence 7, Appli	C 803	15	1.8	2886	9	US-10-180-541-7	Sequence 7, Appli
C 731	15	1.8	2886	9	US-10-175-753-7	Sequence 7, Appli	C 804	15	1.8	2886	9	US-10-180-542-7	Sequence 7, Appli
C 732	15	1.8	2886	9	US-10-176-482-7	Sequence 7, Appli	C 805	15	1.8	2886	9	US-10-180-548-7	Sequence 7, Appli
C 733	15	1.8	2886	9	US-10-176-757-7	Sequence 7, Appli	C 806	15	1.8	2886	9	US-10-180-551-7	Sequence 7, Appli
C 734	15	1.8	2886	9	US-10-176-913-7	Sequence 7, Appli	C 807	15	1.8	2886	9	US-10-180-998-7	Sequence 7, Appli
C 735	15	1.8	2886	9	US-10-180-552-7	Sequence 7, Appli	C 808	15	1.8	2886	9	US-10-180-999-7	Sequence 7, Appli
C 736	15	1.8	2886	9	US-10-180-557-7	Sequence 7, Appli	C 809	15	1.8	2886	9	US-10-183-013-7	Sequence 7, Appli
C 737	15	1.8	2886	9	US-09-990-438-13	Sequence 13, Appli	C 810	15	1.8	2886	9	US-10-184-612-7	Sequence 7, Appli
C 738	15	1.8	2886	9	US-09-990-562-13	Sequence 13, Appli	C 811	15	1.8	2886	9	US-10-184-616-7	Sequence 7, Appli
C 739	15	1.8	2886	9	US-09-997-428-13	Sequence 13, Appli	C 812	15	1.8	2886	9	US-10-184-617-7	Sequence 7, Appli
C 740	15	1.8	2886	9	US-09-997-666-13	Sequence 13, Appli	C 813	15	1.8	2886	9	US-10-184-622-7	Sequence 7, Appli
C 741	15	1.8	2886	9	US-10-173-700-7	Sequence 7, Appli	C 814	15	1.8	2886	9	US-10-184-628-7	Sequence 7, Appli
C 742	15	1.8	2886	9	US-10-174-572-7	Sequence 7, Appli	C 815	15	1.8	2886	9	US-10-184-629-7	Sequence 7, Appli
C 743	15	1.8	2886	9	US-10-174-579-7	Sequence 7, Appli	C 816	15	1.8	2886	9	US-10-184-630-7	Sequence 7, Appli
C 744	15	1.8	2886	9	US-10-174-582-7	Sequence 7, Appli	C 817	15	1.8	2886	9	US-10-184-631-7	Sequence 7, Appli
C 745	15	1.8	2886	9	US-10-174-588-7	Sequence 7, Appli	C 818	15	1.8	2886	9	US-10-184-632-7	Sequence 7, Appli
C 746	15	1.8	2886	9	US-10-175-739-7	Sequence 7, Appli	C 819	15	1.8	2886	9	US-10-184-636-7	Sequence 7, Appli
C 747	15	1.8	2886	9	US-10-175-740-7	Sequence 7, Appli	C 820	15	1.8	2886	9	US-10-184-640-7	Sequence 7, Appli
C 748	15	1.8	2886	9	US-10-175-743-7	Sequence 7, Appli	C 821	15	1.8	2886	9	US-10-184-650-7	Sequence 7, Appli
C 749	15	1.8	2886	9	US-10-176-488-7	Sequence 7, Appli	C 822	15	1.8	2886	9	US-10-184-651-7	Sequence 7, Appli



C 823	15	1.8	2886	9	US-10-187-588-7	Sequence 7, Appli	Sequence 7, Appli	C 896	15	1.8	2886	9	US-10-184-656-7	Sequence 7, Appli
C 824	15	1.8	2886	9	US-10-187-597-7	Sequence 7, Appli	Sequence 7, Appli	C 897	15	1.8	2886	9	US-10-192-010-7	Sequence 7, Appli
C 825	15	1.8	2886	9	US-10-187-598-7	Sequence 7, Appli	Sequence 7, Appli	C 898	15	1.8	2886	9	US-10-195-894-7	Sequence 7, Appli
C 826	15	1.8	2886	9	US-10-187-600-7	Sequence 7, Appli	Sequence 7, Appli	C 899	15	1.8	2886	9	US-10-205-908-7	Sequence 7, Appli
C 827	15	1.8	2886	9	US-10-187-601-7	Sequence 7, Appli	Sequence 7, Appli	C 900	15	1.8	2886	9	US-09-991-157-13	Sequence 13, Appli
C 828	15	1.8	2886	9	US-10-187-602-7	Sequence 7, Appli	Sequence 7, Appli	C 901	15	1.8	2886	9	US-09-991-172-13	Sequence 13, Appli
C 829	15	1.8	2886	9	US-10-187-603-7	Sequence 7, Appli	Sequence 7, Appli	C 902	15	1.8	2886	9	US-09-997-514-13	Sequence 13, Appli
C 830	15	1.8	2886	9	US-10-187-741-7	Sequence 7, Appli	Sequence 7, Appli	C 903	15	1.8	2886	9	US-09-997-514-13	Sequence 13, Appli
C 831	15	1.8	2886	9	US-10-187-746-7	Sequence 7, Appli	Sequence 7, Appli	C 904	15	1.8	2886	9	US-10-186-855-7	Sequence 7, Appli
C 832	15	1.8	2886	9	US-10-187-747-7	Sequence 7, Appli	Sequence 7, Appli	C 905	15	1.8	2886	9	US-10-187-599-7	Sequence 7, Appli
C 833	15	1.8	2886	9	US-10-187-751-7	Sequence 7, Appli	Sequence 7, Appli	C 906	15	1.8	2886	9	US-10-187-750-7	Sequence 7, Appli
C 834	15	1.8	2886	9	US-10-187-753-7	Sequence 7, Appli	Sequence 7, Appli	C 907	15	1.8	2886	9	US-10-188-780-7	Sequence 7, Appli
C 835	15	1.8	2886	9	US-10-187-754-7	Sequence 7, Appli	Sequence 7, Appli	C 908	15	1.8	2886	9	US-10-192-015-7	Sequence 7, Appli
C 836	15	1.8	2886	9	US-10-187-757-7	Sequence 7, Appli	Sequence 7, Appli	C 909	15	1.8	2886	9	US-10-194-394-7	Sequence 7, Appli
C 837	15	1.8	2886	9	US-10-187-758-7	Sequence 7, Appli	Sequence 7, Appli	C 910	15	1.8	2886	9	US-10-194-425-7	Sequence 7, Appli
C 838	15	1.8	2886	9	US-10-187-884-7	Sequence 7, Appli	Sequence 7, Appli	C 911	15	1.8	2886	9	US-10-194-485-7	Sequence 7, Appli
C 839	15	1.8	2886	9	US-10-188-767-7	Sequence 7, Appli	Sequence 7, Appli	C 912	15	1.8	2886	9	US-10-195-885-7	Sequence 7, Appli
C 840	15	1.8	2886	9	US-10-188-769-7	Sequence 7, Appli	Sequence 7, Appli	C 913	15	1.8	2886	9	US-10-195-890-7	Sequence 7, Appli
C 841	15	1.8	2886	9	US-10-188-770-7	Sequence 7, Appli	Sequence 7, Appli	C 914	15	1.8	2886	9	US-10-195-899-7	Sequence 7, Appli
C 842	15	1.8	2886	9	US-10-188-773-7	Sequence 7, Appli	Sequence 7, Appli	C 915	15	1.8	2886	9	US-10-196-748-7	Sequence 7, Appli
C 843	15	1.8	2886	9	US-10-188-781-7	Sequence 7, Appli	Sequence 7, Appli	C 916	15	1.8	2886	9	US-10-196-750-7	Sequence 7, Appli
C 844	15	1.8	2886	9	US-10-194-361-7	Sequence 7, Appli	Sequence 7, Appli	C 917	15	1.8	2886	9	US-10-196-750-7	Sequence 7, Appli
C 845	15	1.8	2886	9	US-10-194-423-7	Sequence 7, Appli	Sequence 7, Appli	C 918	15	1.8	2886	9	US-10-197-699-7	Sequence 7, Appli
C 846	15	1.8	2886	9	US-10-195-897-7	Sequence 7, Appli	Sequence 7, Appli	C 919	15	1.8	2886	9	US-10-197-700-7	Sequence 7, Appli
C 847	15	1.8	2886	9	US-10-195-901-7	Sequence 7, Appli	Sequence 7, Appli	C 920	15	1.8	2886	9	US-10-197-705-7	Sequence 7, Appli
C 848	15	1.8	2886	9	US-10-196-756-7	Sequence 7, Appli	Sequence 7, Appli	C 921	15	1.8	2886	9	US-10-197-708-7	Sequence 7, Appli
C 849	15	1.8	2886	9	US-09-989-726-13	Sequence 13, Appli	Sequence 13, Appli	C 922	15	1.8				

```

c 969      15      1.8      2886      9      US-10-199-462-7      Sequence 7, Appli
c 970      15      1.8      2886      9      US-10-201-324-7      Sequence 7, Appli
c 971      15      1.8      2886      9      US-10-201-327-7      Sequence 7, Appli
c 972      15      1.8      2886      9      US-10-201-527-7      Sequence 7, Appli
c 973      15      1.8      2886      9      US-10-201-528-7      Sequence 7, Appli
c 974      15      1.8      2886      9      US-10-201-529-7      Sequence 7, Appli
c 975      15      1.8      2886      9      US-10-201-530-7      Sequence 7, Appli
c 976      15      1.8      2886      9      US-10-202-408-7      Sequence 7, Appli
c 977      15      1.8      2886      9      US-10-202-411-7      Sequence 7, Appli
c 978      15      1.8      2886      9      US-10-202-472-7      Sequence 7, Appli
c 979      15      1.8      2886      9      US-10-205-502-7      Sequence 7, Appli
c 980      15      1.8      2886      9      US-10-205-507-7      Sequence 7, Appli
c 981      15      1.8      2886      9      US-10-205-517-7      Sequence 7, Appli
c 982      15      1.8      2886      9      US-10-205-511-7      Sequence 7, Appli
c 983      15      1.8      2886      9      US-10-205-902-7      Sequence 7, Appli
c 984      15      1.8      2886      9      US-10-205-907-7      Sequence 7, Appli
c 985      15      1.8      2886      9      US-09-989-729A-13      Sequence 13, Appli
c 986      15      1.8      2886      9      US-09-990-440-13      Sequence 13, Appli
c 987      15      1.8      2886      9      US-09-991-854-13      Sequence 13, Appli
c 988      15      1.8      2886      9      US-09-997-348-13      Sequence 13, Appli
c 989      15      1.8      2886      9      US-09-997-628-13      Sequence 13, Appli
c 990      15      1.8      2886      9      US-09-997-683-13      Sequence 13, Appli
c 991      15      1.8      2886      9      US-10-194-456-7      Sequence 7, Appli
c 992      15      1.8      2886      9      US-10-196-758-7      Sequence 7, Appli
c 993      15      1.8      2886      9      US-10-198-770-7      Sequence 7, Appli
c 994      15      1.8      2886      9      US-10-199-308-7      Sequence 7, Appli
c 995      15      1.8      2886      9      US-10-200-617-7      Sequence 7, Appli
c 996      15      1.8      2886      9      US-10-205-893-7      Sequence 7, Appli
c 997      15      1.8      2886      9      US-10-205-897-7      Sequence 7, Appli
c 998      15      1.8      2886      9      US-09-993-469-13      Sequence 13, Appli
c 999      15      1.8      2886      9      US-09-993-748-13      Sequence 13, Appli
c1000      15      1.8      2886      9      US-09-993-748-13      Sequence 13, Appli

```

## ALIGNMENTS

```

RESULT 1
US-10-062-051-41
; Sequence 41, Application US/10062051
; Publication No. US20030073095A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/10/062,051
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US/09/660,587
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 41
; LENGTH: 840
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: nucleic acid sequence of E. canis p28-2
US-10-062-051-41
Query Match      100.0%; Score 840; DB 9; Length 840;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1      ATGATTATACAAATTCCTAGTACAGCGGTTAATCTCAATGTCATCTTACCA 60
Db      1      ATGATTATACAAATTCCTAGTACAGCGGTTAATCTCAATGTCATCTTACCA 60
Qy      61      TATCAGTCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120

```

FEATURE:  
; OTHER INFORMATION: nucleic acid sequence of E. canis p28-2  
US-10-062-920-41

Query Match 100.0%; Score 840; DB 9; Length 840;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAATTATAAGAAATTTCTAGTAAGAGCGGTTAATCTCATTAAATGTCTTAACCA 60  
Db |||||  
Qy 1 ATGAATTATAAGAAATTTCTAGTAAGAGCGGTTAATCTCATTAAATGTCTTAACCA 60  
Db |||||

Qy 61 TATCAGTCTTTTTCAGATCTCTGAGTTCAGAACTAATGATAACAAGAGGCTTCTAC 120  
Db |||||

Qy 121 ATTAGTCAAAAGTACAAATCCAAAGTATATCAACATTTAGAAATTTCTGCTGAAGAACT 180  
Db |||||

Qy 181 CCTATTAAATGAACAAATTTCTCTCACTAAAGAGTTTTCGGACTAAAGAAAGATGGTGAT 240  
Db |||||

Qy 241 ATACAAAAAAGAGCATTTTACAGAGTAGTCTCAGGCATTTGATTTTCAAAATAACTTA 300  
Db |||||

Qy 241 ATACAAAAAAGAGCATTTTACAGAGTAGTCTCAGGCATTTGATTTTCAAAATAACTTA 300  
Db |||||

Qy 301 ATATCAGATTTTCAGGAAGTATTTGTTTACCTATGAGCGGACCAAGAAATAGAACTTGA 360  
Db |||||

Qy 301 ATATCAGATTTTCAGGAAGTATTTGTTTACCTATGAGCGGACCAAGAAATAGAACTTGA 360  
Db |||||

Qy 361 GCTGCATATCAACAAATTTAATCCAAAAACCCGATATGATGATCTGATTAATGGTGAA 420  
Db |||||

Qy 421 TACTATAAATTTTGCATTTCTCGTAAGATGCAATGCAATGATGATGATGATGATGATGAT 480  
Db |||||

Qy 481 CTTAAAAATGAGCGATTAACCTTTTATGTTGATGATGATGATGATGATGATGATGATGAT 540  
Db |||||

Qy 481 CTTAAAAATGAGCGATTAACCTTTTATGTTGATGATGATGATGATGATGATGATGATGAT 540  
Db |||||

Qy 541 GCTGAAGAGTATCTTTTCGTAACCATATGATGATGATGATGATGATGATGATGATGATGAT 600  
Db |||||

Qy 601 ATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGAAAAATAGGATTTAGTTACCT 660  
Db |||||

Qy 661 ATACACAGAGTCTCTGCAATTTATGTTGATGATGATGATGATGATGATGATGATGATGAT 720  
Db |||||

Qy 661 ATACACAGAGTCTCTGCAATTTATGTTGATGATGATGATGATGATGATGATGATGATGAT 720  
Db |||||

Qy 721 TTTGAGAAGATACCTGTAATAACTCTCTGATGATGATGATGATGATGATGATGATGATGAT 780  
Db |||||

Qy 721 TTTGAGAAGATACCTGTAATAACTCTCTGATGATGATGATGATGATGATGATGATGATGAT 780  
Db |||||

Qy 781 GCTTCAGTAACTCTTGAGCTTGGATATCTTTGGCGGAGAAATTTGGAATGAGGTTTACCTTC 840  
Db |||||

## RESULT 3

US-10-062-624-41  
; Sequence 41, Application US/10062624  
; Patent No. US20020115840A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie

; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof  
; FILE REFERENCE: D6152CIP2/D1  
; CURRENT APPLICATION NUMBER: US/10/062,624  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: 09/660,587  
; PRIOR FILING DATE: 2000-09-12  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 41  
; LENGTH: 840  
; TYPE: DNA  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: nucleic acid sequence of E. canis p28-2  
US-10-062-624-41

Query Match 100.0%; Score 840; DB 12; Length 840;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAATTATAAGAAATTTCTAGTAAGAGCGGTTAATCTCATTAAATGTCTTAACCA 60  
Db |||||

Qy 61 TATCAGTCTTTTTCAGATCTCTGAGTTCAGAACTAATGATAACAAGAGGCTTCTAC 120  
Db |||||

Qy 121 ATTAGTCAAAAGTACAAATCCAAAGTATATCAACATTTAGAAATTTCTGCTGAAGAACT 180  
Db |||||

Qy 181 CCTATTAAATGAACAAATTTCTCTCACTAAAAAGTTTTCGGACTAAAGAAAGATGGTGAT 240  
Db |||||

Qy 241 ATACAAAAAAGAGCATTTTACAGAGTAGTCTCAGGCATTTGATTTTCAAAATAACTTA 300  
Db |||||

Qy 301 ATATCAGATTTTCAGGAAGTATTTGTTTACCTATGAGCGGACCAAGAAATAGAACTTGA 360  
Db |||||

Qy 361 GCTGCATATCAACAAATTTAATCCAAAAACCCGATATGATGATGATGATGATGATGATGAT 420  
Db |||||

Qy 421 TACTATAAATTTTGCATTTCTCGTAAGATGCAATGCAATGATGATGATGATGATGATGAT 480  
Db |||||

Qy 481 CTTAAAAATGAGCGATTAACCTTTTATGTTGATGATGATGATGATGATGATGATGATGAT 540  
Db |||||

Qy 481 CTTAAAAATGAGCGATTAACCTTTTATGTTGATGATGATGATGATGATGATGATGATGAT 540  
Db |||||

Qy 541 GCTGAAGAGTATCTTTTCGTAACCATATGATGATGATGATGATGATGATGATGATGATGAT 600  
Db |||||

Qy 601 ATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGAAAAATAGGATTTAGTTACCT 660  
Db |||||

Qy 661 ATACACAGAGTCTCTGCAATTTATGTTGATGATGATGATGATGATGATGATGATGATGAT 720  
Db |||||

Qy 661 ATACACAGAGTCTCTGCAATTTATGTTGATGATGATGATGATGATGATGATGATGATGAT 720  
Db |||||

Qy 721 TTTGAGAAGATACCTGTAATAACTCTCTGATGATGATGATGATGATGATGATGATGATGAT 780  
Db |||||

Qy 721 TTTGAGAAGATACCTGTAATAACTCTCTGATGATGATGATGATGATGATGATGATGATGAT 780  
Db |||||

Qy 781 GCTTCAGTAACTCTTGAGCTTGGATATCTTTGGCGGAGAAATTTGGAATGAGGTTTACCTTC 840  
Db |||||

```
Db      781 GCTTCAGTAACCTCTTGACGTTGGATACCTTTGGCGGAGAAATGGATAGGTTTCACTTC 840
|||||
RESULT 4
US-10-314-639-47
; Sequence 47, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/314,701
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 843
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(843)
US-10-314-639-47

Query Match      100.0%; Score 840; DB 9; Length 843;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATGAATTATAAGAAATCTAGTAAGACGCGTTAACTCTCAATTAATGTCATCTTACCA 60
DB      1 ATGAATTATAAGAAATCTAGTAAGACGCGTTAACTCTCAATTAATGTCATCTTACCA 60

QY      61 TATCAGTCTTTGCGAGCTCTAGTTCAGAACTATATGATTAACAAGAGGCTTCTAC 120
DB      61 TATCAGTCTTTGCGAGCTCTAGTTCAGAACTATATGATTAACAAGAGGCTTCTAC 120

QY      121 ATTAGTGAACAAATCTCTCACTAAAGAAATTTTCGACTAAAGAAAGATGGTGAT 180
DB      121 ATTAGTGAACAAATCTCTCACTAAAGAAATTTTCGACTAAAGAAAGATGGTGAT 180

QY      181 CCTATTATGGAACAAATCTCTCACTAAAGAAATTTTCGACTAAAGAAAGATGGTGAT 240
DB      181 CCTATTATGGAACAAATCTCTCACTAAAGAAATTTTCGACTAAAGAAAGATGGTGAT 240

QY      241 ATACAAAAAAGACGATTTTACAGAGTAGCTCCAGGATTTGATTTTCAAAATAACTTA 300
DB      241 ATACAAAAAAGACGATTTTACAGAGTAGCTCCAGGATTTGATTTTCAAAATAACTTA 300

QY      301 ATATCAGATTTTCAGGAAGTATTTGCTACTTATGATGGTTAACTTGTCTATGACATTACA 360
DB      301 ATATCAGATTTTCAGGAAGTATTTGCTACTTATGATGGTTAACTTGTCTATGACATTACA 360

QY      361 GCTGATATCAACAAATTTTAAATCCAAAAACACCGATAACATGATGATGATGATGATGAT 420
DB      361 GCTGATATCAACAAATTTTAAATCCAAAAACACCGATAACATGATGATGATGATGATGAT 420

QY      421 TACTATAACATTTTGCATATCTCTGTAAGATGCAATGGAAGATCAGCAATATGTAGTA 480
DB      421 TACTATAACATTTTGCATATCTCTGTAAGATGCAATGGAAGATCAGCAATATGTAGTA 480

QY      481 CTTAAAAATGACGGATACCTTTTATGCTATGATGGTTAACTTGTCTATGACATTACA 540
DB      481 CTTAAAAATGACGGATACCTTTTATGCTATGATGGTTAACTTGTCTATGACATTACA 540

QY      541 GCTGAAGGATATCTTTTGTGATACATATGATGATGATGATGATGATGATGATGATGATGAT 600
DB      541 GCTGAAGGATATCTTTTGTGATACATATGATGATGATGATGATGATGATGATGATGATGAT 600
```

```
QY      601 ATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGAAAAATAGGTATTAGTTACCT 660
DB      601 ATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGAAAAATAGGTATTAGTTACCT 660

QY      661 ATCACACCAGAACTCTCTGCAATTTTATTTGGTGGATATACCATGGCGTTATTGGTAATAA 720
DB      661 ATCACACCAGAACTCTCTGCAATTTTATTTGGTGGATATACCATGGCGTTATTGGTAATAA 720

QY      721 TTTGAGAAGATACCTGTAAATTAATCTCTGTAGTATTAATGATGCTCTCAAAACCATCT 780
DB      721 TTTGAGAAGATACCTGTAAATTAATCTCTGTAGTATTAATGATGCTCTCAAAACCATCT 780

QY      781 GCTTCAGTAACCTCTTGAGCTTTGGATATCTTTCGCGAGAAATTTGGAATGAGGTTCACTTC 840
DB      781 GCTTCAGTAACCTCTTGAGCTTTGGATATCTTTCGCGAGAAATTTGGAATGAGGTTCACTTC 840

RESULT 5
US-10-059-964-47
; Sequence 47, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; CURRENT FILING DATE: 2002-01-28
; EARLIER APPLICATION NUMBER: 09/314,701
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 843
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(843)
US-10-059-964-47

Query Match      100.0%; Score 840; DB 12; Length 843;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATGAATTATAAGAAATCTAGTAAGACGCGTTAACTCTCAATTAATGTCATCTTACCA 60
DB      1 ATGAATTATAAGAAATCTAGTAAGACGCGTTAACTCTCAATTAATGTCATCTTACCA 60

QY      61 TATCAGTCTTTGCGAGCTCTAGTTCAGAACTATATGATTAACAAGAGGCTTCTAC 120
DB      61 TATCAGTCTTTGCGAGCTCTAGTTCAGAACTATATGATTAACAAGAGGCTTCTAC 120

QY      121 ATTAGTGAACAAATCTCTCACTAAAGAAATTTTCGACTAAAGAAAGATGGTGAT 180
DB      121 ATTAGTGAACAAATCTCTCACTAAAGAAATTTTCGACTAAAGAAAGATGGTGAT 180

QY      181 CCTATTATGGAACAAATCTCTCACTAAAGAAATTTTCGACTAAAGAAAGATGGTGAT 240
DB      181 CCTATTATGGAACAAATCTCTCACTAAAGAAATTTTCGACTAAAGAAAGATGGTGAT 240

QY      241 ATACAAAAAAGACGATTTTACAGAGTAGCTCCAGGATTTGATTTTCAAAATAACTTA 300
DB      241 ATACAAAAAAGACGATTTTACAGAGTAGCTCCAGGATTTGATTTTCAAAATAACTTA 300

QY      301 ATATCAGATTTTCAGGAAGTATTTGCTACTTATGATGGTTAACTTGTCTATGACATTACA 360
DB      301 ATATCAGATTTTCAGGAAGTATTTGCTACTTATGATGGTTAACTTGTCTATGACATTACA 360

QY      361 GCTGATATCAACAAATTTTAAATCCAAAAACACCGATAACATGATGATGATGATGATGAT 420
DB      361 GCTGATATCAACAAATTTTAAATCCAAAAACACCGATAACATGATGATGATGATGATGAT 420
```



; PRIOR APPLICATION NUMBER: US/09/660,587  
; PRIOR FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 37  
; LENGTH: 22  
; TYPE: DNA  
; ORGANISM: artificial sequence  
; FEATURE:  
; OTHER INFORMATION: primer ECa5'-1600  
US-10-062-920-37

Query Match 2.6%; Score 22; DB 9; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.8;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 670 GAAGTCTCTGCATTTATTGGTG 691  
|||||  
DB 22 GAAGTCTCTGCATTTATTGGTG 1

## RESULT 10

US-10-062-624-37/c  
; Sequence 37, Application US/10062624  
; Publication No. US20020115840A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-Kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2/D1  
; CURRENT APPLICATION NUMBER: US/10/062,624  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: 09/660,587  
; PRIOR FILING DATE: 2000-09-12  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 37  
; LENGTH: 22  
; TYPE: DNA  
; ORGANISM: artificial sequence  
; FEATURE:  
; OTHER INFORMATION: primer ECa5'-1600  
US-10-062-624-37

Query Match 2.6%; Score 22; DB 12; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.8;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 670 GAAGTCTCTGCATTTATTGGTG 691  
|||||  
DB 22 GAAGTCTCTGCATTTATTGGTG 1

## RESULT 11

US-10-062-051-5  
; Sequence 5, Application US/10062051  
; Publication No. US20030073095A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: US/10/062,051  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: US/09/660,587  
; PRIOR FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46

; SEQ ID NO 5  
; LENGTH: 840  
; TYPE: DNA  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; NAME/KEY: mat peptide  
; OTHER INFORMATION: nucleic acid sequence of p28-6  
US-10-062-051-5

Query Match 2.3%; Score 19; DB 9; Length 840;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 343 CCAAGAATAGAACTTGAAG 361  
|||||  
DB 337 CCAAGAATAGAACTTGAAG 355

## RESULT 12

US-10-062-920-5  
; Sequence 5, Application US/10062920  
; Publication No. US20030096250A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: US/10/062,920  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: US/09/660,587  
; PRIOR FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 5  
; LENGTH: 840  
; TYPE: DNA  
; ORGANISM: Ehrlichia canis  
; NAME/KEY: mat peptide  
; OTHER INFORMATION: nucleic acid sequence of p28-6  
US-10-062-920-5

Query Match 2.3%; Score 19; DB 9; Length 840;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 343 CCAAGAATAGAACTTGAAG 361  
|||||  
DB 337 CCAAGAATAGAACTTGAAG 355

## RESULT 13

US-10-062-624-5  
; Sequence 5, Application US/10062624  
; Patent No. US20020115840A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-Kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2/D1  
; CURRENT APPLICATION NUMBER: US/10/062,624  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: 09/660,587  
; PRIOR FILING DATE: 2000-09-12  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 5  
; LENGTH: 840  
; TYPE: DNA

```
; ORGANISM: Ehrlichia canis
; FEATURE:
; NAME/KEY: mat_peptide
; OTHER INFORMATION: nucleic acid sequence of p28-6
US-10-062-624-5

Query Match      2.3%; Score 19; DB 12; Length 840;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 343 CCAAGAATAGAACTTGAAG 361
Db 337 CCAAGAATAGAACTTGAAG 355

RESULT 14
US-10-314-639-37
; Sequence 37, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; PRIOR FILING DATE: 2002-12-09
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 843
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(843)
US-10-314-639-37

Query Match      2.3%; Score 19; DB 9; Length 843;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 343 CCAAGAATAGAACTTGAAG 361
Db 337 CCAAGAATAGAACTTGAAG 355

RESULT 15
US-10-059-964-37
; Sequence 37, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; CURRENT FILING DATE: 2002-01-28
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 843
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(843)
US-10-059-964-37

Query Match      2.3%; Score 19; DB 12; Length 840;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 343 CCAAGAATAGAACTTGAAG 361
Db 337 CCAAGAATAGAACTTGAAG 355

RESULT 16
US-10-314-639-13
; Sequence 13, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; CURRENT FILING DATE: 2002-12-09
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 894
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(894)
US-10-314-639-13

Query Match      2.3%; Score 19; DB 9; Length 894;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 620 TAAATTTTGCTTACCAAGG 638
Db 671 TAAATTTTGCTTACCAAGG 689

RESULT 17
US-10-059-964-13
; Sequence 13, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; CURRENT FILING DATE: 2002-01-28
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 894
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(894)
US-10-059-964-13

Query Match      2.3%; Score 19; DB 12; Length 894;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 620 TAAATTTGCTTACCAAG 638  
Db 671 TAAATTTGCTTACCAAG 699

## RESULT 18

US-10-239-676-18/c  
; Sequence 18, Application US/10239676  
; Publication No. US20030082609A1  
; GENERAL INFORMATION:  
; APPLICANT: PIEPNEROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
; FILE REFERENCE: 5013.1003  
; CURRENT APPLICATION NUMBER: US/10/239,676  
; CURRENT FILING DATE: 2002-09-24  
; PRIOR APPLICATION NUMBER: PCT/EP01/03968  
; DE 10019058.8  
; DE 10019173.8  
; DE 10032529.7  
; DE 10043826.1  
; PRIOR FILING DATE: 2001-04-06  
; 2000-04-06  
; 2000-04-07  
; 2000-06-30  
; 2000-09-01  
; NUMBER OF SEQ ID NOS: 228  
; SEQ ID NO 18  
; LENGTH: 5979  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-239-676-18

Query Match 2.3%; Score 19; DB 9; Length 5979;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 TTTTCAAAATACTTAATA 303  
Db 1203 TTTTCAAAATACTTAATA 1185

## RESULT 19

US-09-864-761-1982/c  
; Sequence 1982, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aomicra-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 1982  
; LENGTH: 429  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC011242.1  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.1  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4  
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 2.9  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.1  
; OTHER INFORMATION: EXPRESSED IN HEAT, SIGNAL = 3.6  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.4  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.3  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.7  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.6  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.4  
US-09-864-761-1982

Query Match 2.1%; Score 18; DB 10; Length 429;  
Best Local Similarity 100.0%; Pred. No. 94;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 AAGGAAAAATAGGTATTA 652  
Db 131 AAGGAAAAATAGGTATTA 114

## RESULT 20

US-09-770-961-540  
; Sequence 540, Application US/09770961  
; Publication No. US20030115639A1  
; GENERAL INFORMATION:  
; APPLICANT: Goriach, Jorn  
; APPLICANT: An, Yong-Qiang  
; APPLICANT: Hamilton, Carol M.  
; APPLICANT: Price, Jennifer L.  
; APPLICANT: Raines, Tracy M.  
; APPLICANT: Yu, Yang  
; APPLICANT: Rameaka, Joshua G.  
; APPLICANT: Page, Amy  
; APPLICANT: Matthew, Abraham V.  
; APPLICANT: Ledford, Brooke L.  
; APPLICANT: Woessner, Jeffrey P.  
; APPLICANT: Haas, William David  
; APPLICANT: Garcia, Carlos A.  
; APPLICANT: Kricker, Maja  
; APPLICANT: Slader, Ted  
; APPLICANT: Davis, Keith R.  
; APPLICANT: Allen, Keith



; APPLICANT: Hoffman, Neil  
; APPLICANT: Hurlan, Patrick  
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
; TITLE OF INVENTION: thaliana  
; FILE REFERENCE: 2026 (PARA-015PRV)  
; CURRENT APPLICATION NUMBER: US/09/770,961  
; CURRENT FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: 60/178,466  
; PRIOR FILING DATE: 2000-01-27  
; NUMBER OF SEQ ID NOS: 999  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 540  
; LENGTH: 495  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-770-961-540

Query Match 2.1%; Score 18; DB 9; Length 495;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 808 TTGGCGGAGAAATTGGA 825  
|||  
Db 358 TTGGCGGAGAAATTGGA 375

RESULT 21  
US-09-864-761-8939  
; Sequence 8939, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aecmca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 8939  
; LENGTH: 588  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC018573.2  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.9  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.3  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.5  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.6  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.6  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.1  
US-09-864-761-8939

Query Match 2.1%; Score 18; DB 10; Length 588;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 461 AAGATCAGCAATATGTAG 478  
|||  
Db 357 AAGATCAGCAATATGTAG 374

RESULT 22  
US-09-938-842A-1696  
; Sequence 1696, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Zhu, Tong  
; APPLICANT: Wang, Xun  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; TITLE OF INVENTION: SAME, AND METHODS OF USE  
; FILE REFERENCE: SCRIPL300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; CURRENT FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 1696  
; LENGTH: 933  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-1696

Query Match 2.1%; Score 18; DB 9; Length 933;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 808 TTGGCGGAGAAATTGGA 825  
|||  
Db 249 TTGGCGGAGAAATTGGA 266

RESULT 23  
US-09-323-998D-3  
; Sequence 3, Application US/09323998D  
; Patent No. US20020102631A1  
; GENERAL INFORMATION:  
; APPLICANT: CUNNINGHAM JR., FRANCIS X.  
; APPLICANT: SUN, ZAIREN

;; TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND  
;; TITLE OF INVENTION: METHODS OF USE THEREOF  
;; FILE REFERENCE: 108172-09019  
;; CURRENT APPLICATION NUMBER: US/09/323,998D  
;; CURRENT FILING DATE: 1999-06-02  
;; PRIOR APPLICATION NUMBER: 09/088,724  
;; PRIOR FILING DATE: 1998-06-02  
;; PRIOR APPLICATION NUMBER: 09/088,725  
;; PRIOR FILING DATE: 1998-06-02  
;; PRIOR APPLICATION NUMBER: 08/937,155  
;; PRIOR FILING DATE: 1997-09-25  
;; PRIOR APPLICATION NUMBER: 08/624,125  
;; PRIOR FILING DATE: 1996-03-29  
;; NUMBER OF SEQ ID NOS: 61  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 3  
;; LENGTH: 956  
;; TYPE: DNA  
;; ORGANISM: Arabidopsis thaliana  
US-09-323-998D-3

Query Match 2.1%; Score 18; DB 10; Length 956;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 808 TTGGCGGAGAAATTGGA 825  
DB 203 TTGGCGGAGAAATTGGA 220

RESULT 24  
US-10-001-843-25  
; Sequence 25, Application US/10001843  
; Patent No. US20020132255A1  
; GENERAL INFORMATION:  
; APPLICANT: Salceda, Susana  
; APPLICANT: Macina, Roberto  
; APPLICANT: Recipon, Herve  
; APPLICANT: Cafferkey, Robert  
; APPLICANT: Sun, Yongming  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Turner, Leah  
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P  
; FILE REFERENCE: DEX-0267  
; CURRENT APPLICATION NUMBER: US/10/001,843  
; CURRENT FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: 60/249,992  
; PRIOR FILING DATE: 2000-11-20  
; NUMBER OF SEQ ID NOS: 218  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 25  
; LENGTH: 1301  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (520)..(520)  
; OTHER INFORMATION: a, c, g or t  
US-10-001-843-25

Query Match 2.1%; Score 18; DB 12; Length 1301;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GAATTATAGAAAATCT 20  
DB 303 GAATTATAGAAAATCT 320

RESULT 25  
US-08-781-986A-551  
; Sequence 551, Application US/08781986A  
; Publication No. US20030054436A1

;; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5255  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781,986A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Renson, Bob  
; REGISTRATION NUMBER: 30,446  
; REFERENCE/DOCKET NUMBER: PB248PP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 551:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4923 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-781-986A-551

Query Match 2.1%; Score 18; DB 7; Length 4923;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 AAGAAAGATGGTGATATA 243  
DB 4355 AAGAAAGATGGTGATATA 4372

RESULT 26  
US-10-210-296-1  
; Sequence 1, Application US/10210296  
; Publication No. US20030021802A1  
; GENERAL INFORMATION:  
; APPLICANT: Pfizer Products Inc.  
; TITLE OF INVENTION: LAWSONIA INTRACELLULARIS PROTEINS, AND RELATED  
; TITLE OF INVENTION: METHODS  
; TITLE OF INVENTION: AND MATERIALS  
; FILE REFERENCE: PC10589A  
; CURRENT APPLICATION NUMBER: US/10/210,296  
; CURRENT FILING DATE: 2002-08-01  
; PRIOR APPLICATION NUMBER: US/09/689,065  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 102  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 6617  
; TYPE: DNA  
; ORGANISM: Lawsonia intracellularis  
US-10-210-296-1

Query Match 2.1%; Score 18; DB 9; Length 6617;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 AATCTCTGCTGAAGAAA 178  
|||||  
Db 3326 AATCTCTGCTGAAGAAA 3343

RESULT 27  
US-10-095-407-16  
; Sequence 16, Application US/10095407  
; Patent No. US20020164330A1  
; GENERAL INFORMATION:  
; APPLICANT: Pan, Yang  
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY  
; FILE REFERENCE: 09404/052001  
; CURRENT APPLICATION NUMBER: US/10/095,407  
; CURRENT FILING DATE: 2002-03-11  
; PRIOR APPLICATION NUMBER: US 60/091,650  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: US 60/054,646  
; PRIOR FILING DATE: 1997-08-04  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 16  
; LENGTH: 152331  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(152331)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-095-407-16

Query Match 2.1%; Score 18; DB 9; Length 152331;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AATTATAAGAAATTTCTA 21  
|||||  
Db 61113 AATTATAAGAAATTTCTA 61130

RESULT 28  
US-09-835-232-7/c  
; Sequence 7, Application US/09835232  
; Patent No. US20020098489A1  
; GENERAL INFORMATION:  
; APPLICANT: Leder, Philip  
; TITLE OF INVENTION: FORMIN-2 NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 00383/052002  
; CURRENT APPLICATION NUMBER: US/09/835,232  
; CURRENT FILING DATE: 2001-04-12  
; PRIOR APPLICATION NUMBER: US 60/196,811  
; PRIOR FILING DATE: 2000-04-13  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 170834  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(170834)  
; OTHER INFORMATION: n = A,T,C, or G  
US-09-835-232-7

Query Match 2.1%; Score 18; DB 10; Length 170834;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 TATAAACATTTTGCATTA 441  
|||||

Db 84231 TATAAACATTTTGCATTA 84214

RESULT 29  
US-10-095-407-17/c  
; Sequence 17, Application US/10095407  
; Patent No. US20020164330A1  
; GENERAL INFORMATION:  
; APPLICANT: Pan, Yang  
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY  
; FILE REFERENCE: 09404/052001  
; CURRENT APPLICATION NUMBER: US/10/095,407  
; CURRENT FILING DATE: 2002-03-11  
; PRIOR APPLICATION NUMBER: US 60/091,650  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: US 60/054,646  
; PRIOR FILING DATE: 1997-08-04  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 17  
; LENGTH: 176373  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(176373)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-095-407-17

Query Match 2.1%; Score 18; DB 9; Length 176373;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AATTATAAGAAATTTCTA 21  
|||||  
Db 110497 AATTATAAGAAATTTCTA 110480

RESULT 30  
US-10-020-141-5  
; Sequence 5, Application US/10020141  
; Publication No. US20030092013A1  
; GENERAL INFORMATION:  
; APPLICANT: McCarthy, Jeanette  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF VASCULAR DISEASE  
; FILE REFERENCE: MMI-002  
; CURRENT APPLICATION NUMBER: US/10/020,141  
; CURRENT FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: US 60/313,097  
; PRIOR FILING DATE: 2001-08-16  
; PRIOR APPLICATION NUMBER: US 60/327,485  
; PRIOR FILING DATE: 2001-10-05  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 183337  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-020-141-5

Query Match 2.1%; Score 18; DB 9; Length 183337;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 AGAAATTTCTGCTGAA 174  
|||||  
Db 7385 AGAAATTTCTGCTGAA 7402

RESULT 31  
US-09-949-654-3

; Sequence 3, Application US/09949654  
; Patent No. US20020127644A1  
; GENERAL INFORMATION:  
; APPLICANT: YAN, Chunhua et al.  
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: CL000817  
; CURRENT APPLICATION NUMBER: US/09/949,654  
; CURRENT FILING DATE: 2001-09-12  
; PRIOR APPLICATION NUMBER: 60/231,572  
; PRIOR FILING DATE: 2000-09-11  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 368004  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(368004)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-654-3

Query Match 2.1%; Score 18; DB 10; Length 368004;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 AACTAATGATACAAAGA 110  
|||||  
DB 119785 AACTAATGATACAAAGA 119802

RESULT 32  
US-09-345-373-35/c  
; Sequence 35, Application US/09345373  
; Publication No. US2003007695A1  
; GENERAL INFORMATION:  
; APPLICANT: RUBEN, STEVEN M.  
; APPLICANT: JIMENEZ, PABLO  
; APPLICANT: DUAN, D. ROXANNE  
; APPLICANT: RAMPY, MARK A.  
; APPLICANT: MENDRICK, DONNA  
; APPLICANT: ZHANG, JUN  
; APPLICANT: NI, JIAN  
; APPLICANT: MOORE, PAUL A.  
; APPLICANT: COLEMAN, TIMOTHY A.  
; APPLICANT: GRUBER, JOACHIM R.  
; APPLICANT: DILLON, PATRICK J.  
; APPLICANT: GENTZ, REINER L.  
; TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2  
; NUMBER OF SEQUENCES: 148  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
; STREET: 1100 NEW YORK AVE, NW, SUITE 600  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/345,373  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/023,082  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/461,195

; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/023,852  
; FILING DATE: 13-AUG-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/039,045  
; FILING DATE: 28-FEB-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/862,432  
; FILING DATE: 23-MAY-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/910,875  
; FILING DATE: 13-AUG-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/055,561  
; FILING DATE: 13-AUG-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEFFEE, ERIC K.  
; REGISTRATION NUMBER: 36,688  
; REFERENCE/DOCKET NUMBER: 1488.0360008/EKS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-2600  
; TELEFAX: 202-371-2540  
; INFORMATION FOR SEQ ID NO: 35:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 75 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-09-345-373-35

Query Match 2.0%; Score 17; DB 9; Length 75;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 GATTTTCAGGAAGTATT 324  
|||||  
DB 45 GATTTTCAGGAAGTATT 29

RESULT 33  
US-10-194-443-19/c  
; Sequence 19, Application US/10194443  
; Publication No. US20030119108A1  
; GENERAL INFORMATION:  
; APPLICANT: Laird, Michael W.  
; TITLE OF INVENTION: Method of Making FGF-12 and KGF-2  
; FILE REFERENCE: 1488.1690001  
; CURRENT APPLICATION NUMBER: US/10/194,443  
; CURRENT FILING DATE: 2002-07-12  
; PRIOR APPLICATION NUMBER: US 60/304,642  
; PRIOR FILING DATE: 2001-07-12  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 19  
; LENGTH: 75  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Oligonucleotide primer  
US-10-194-443-19

Query Match 2.0%; Score 17; DB 9; Length 75;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 GATTTTCAGGAAGTATT 324  
|||||  
DB 45 GATTTTCAGGAAGTATT 29

RESULT 34

US-09-345-373-103/c  
; Sequence 103, Application US/09345373  
; Publication No. US20030077695A1  
; GENERAL INFORMATION:  
; APPLICANT: RUBEN, STEVEN M.  
; APPLICANT: JIMENEZ, PABLO  
; APPLICANT: DUAN, D. ROXANNE  
; APPLICANT: RAMPEY, MARK A.  
; APPLICANT: MENDRICK, DONNA  
; APPLICANT: ZHANG, JUN  
; APPLICANT: NI, JIAN  
; APPLICANT: MOORE, PAUL A.  
; APPLICANT: COLEMAN, TIMOTHY A.  
; APPLICANT: GRUBER, JOACHIM R.  
; APPLICANT: DILLON, PATRICK J.  
; APPLICANT: GENTZ, REINER L.  
; TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2  
; NUMBER OF SEQUENCES: 148  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
; STREET: 1100 NEW YORK AVE, NW, SUITE 600  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/345,373  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/023,082  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/461,195  
; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/023,852  
; FILING DATE: 13-AUG-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/039,045  
; FILING DATE: 28-FEB-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/862,432  
; FILING DATE: 23-MAY-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/910,875  
; FILING DATE: 13-AUG-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/055,561  
; FILING DATE: 13-AUG-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEFFKE, ERIC K.  
; REGISTRATION NUMBER: 36,688  
; REFERENCE/DOCKET NUMBER: 1488.0360008/EKS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-2600  
; TELEFAX: 202-371-2540  
; INFORMATION FOR SEQ ID NO: 103:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 90 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; US-09-345-373-103

Query Match 2.0%; Score 17; DB 9; Length 90;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 308 GATTTTCAGGAAGTATT 324  
|||||  
Db 60 GATTTTCAGGAAGTATT 44  
|||||  
RESULT 35  
US-09-853-666-24/c  
; Sequence 24, Application US/09853666  
; Patent No. US20020016295A1  
; GENERAL INFORMATION:  
; APPLICANT: Gentz, Reiner L.  
; APPLICANT: Chopra, Arvind  
; APPLICANT: Kaushal, Parveen  
; APPLICANT: Spitznagel, Thomas  
; APPLICANT: Unsworth, Edward  
; APPLICANT: Khan, Fazal  
; TITLE OF INVENTION: Keratinocyte Growth Factor-2 Formulations  
; FILE REFERENCE: 1488.1030001  
; CURRENT APPLICATION NUMBER: US/09/853,666  
; CURRENT FILING DATE: 2001-05-14  
; PRIOR APPLICATION NUMBER: 09/218,444  
; PRIOR FILING DATE: 1998-12-22  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 90  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-853-666-24

Query Match 2.0%; Score 17; DB 10; Length 90;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 308 GATTTTCAGGAAGTATT 324  
|||||  
Db 60 GATTTTCAGGAAGTATT 44  
|||||

## RESULT 36

US-09-878-574-10764  
; Sequence 10764, Application US/09878574  
; Patent No. US20020110548A1  
; GENERAL INFORMATION:  
; APPLICANT: Byrum, Joseph R.  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Thompson, Michael D.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(15401)B  
; CURRENT APPLICATION NUMBER: US/09/878,574  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/333,535  
; PRIOR FILING DATE: 1999-06-14  
; NUMBER OF SEQ ID NOS: 15775  
; SEQ ID NO 10764  
; LENGTH: 256  
; TYPE: DNA  
; ORGANISM: Glycine max  
; OTHER INFORMATION: Clone ID: 700968039H1  
US-09-878-574-10764

Query Match 2.0%; Score 17; DB 10; Length 256;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 198 TTCTCTCACTAAAAAAG 214  
|||||  
Db 52 TTCTCTCACTAAAAAAG 68  
|||||

RESULT 37  
US-09-345-373-81/c  
; Sequence 81, Application US/09345373  
; Publication No. US20030077695A1  
; GENERAL INFORMATION:  
; APPLICANT: RUBEN, STEVEN M.  
; APPLICANT: JIMENEZ, PABLO  
; APPLICANT: DUAN, D. ROXANNE  
; APPLICANT: RAMPY, MARK A.  
; APPLICANT: MENDRICK, DONNA  
; APPLICANT: ZHANG, JUN  
; APPLICANT: NI, JIAN  
; APPLICANT: MOORE, PAUL A.  
; APPLICANT: COLEMAN, TIMOTHY A.  
; APPLICANT: GRUBER, JOACHIM R.  
; APPLICANT: DILLON, PATRICK J.  
; APPLICANT: GENTZ, REINER L.  
; TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2  
; NUMBER OF SEQUENCES: 148  
; CORRESPONDENCE ADDRESS: 148  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
; STREET: 1100 NEW YORK AVE, NW, SUITE 600  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/345,373  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/023,082  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/461,195  
; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/023,852  
; FILING DATE: 13-AUG-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/039,045  
; FILING DATE: 28-FEB-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/862,432  
; FILING DATE: 23-MAY-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/910,875  
; FILING DATE: 13-AUG-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/055,561  
; FILING DATE: 13-AUG-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEFFE, ERIC K.  
; REGISTRATION NUMBER: 36,688  
; REFERENCE/DOCKET NUMBER: 1488.0360008/EKS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-2600  
; TELEFAX: 202-371-2540  
; INFORMATION FOR SEQ ID NO: 81:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 276 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: both  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..276

US-09-345-373-81  
Query Match 2.0%; Score 17; DB 9; Length 276;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 308 GATTTTCAGGAAGTATT 324  
Db 93 GATTTTCAGGAAGTATT 77  
RESULT 38  
US-10-091-572-579  
; Sequence 579, Application US/10091572  
; Publication No. US20030054373A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PAl18C1  
; CURRENT APPLICATION NUMBER: US/10/091,572  
; CURRENT FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 09/764,850  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/214,886  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 60/217,487  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,758  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,963  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/217,496  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,447  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/219,290  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/225,757  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/226,868  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/216,647  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: 60/225,267  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/216,880  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: 60/225,270  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/251,869  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/235,834  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/234,274  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: 60/234,223  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: 60/228,924  
; PRIOR FILING DATE: 2000-08-30  
; PRIOR APPLICATION NUMBER: 60/224,518  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/236,369  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/224,519  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,964  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/241,809  
; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/249,299  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/236,327  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/241,785  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/244,617  
; PRIOR FILING DATE: 2000-11-01  
; PRIOR APPLICATION NUMBER: 60/225,268  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/236,368  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/251,856  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/251,868  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/229,344  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/234,997  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: 60/229,343  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,345  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,287  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,513  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: 60/231,413  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/229,509  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: 60/236,367  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/237,039  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,038  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/236,370  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/236,802  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,037  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,040  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/240,960  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/239,935  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: 60/239,937  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: 60/241,787  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/246,474  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/246,532  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/249,216  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,210  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/226,681  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/225,759  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/225,213  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/227,182  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/225,214  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/235,836

; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/230,438  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/215,135  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: 60/225,266  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/249,218  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,208  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,213  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,212  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,207  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,245  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,244  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,217  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,211  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,215  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,264  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/232,400  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/231,242  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/232,081  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/232,080  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/231,414  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/231,244  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/233,064  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/233,063  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/232,397  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/232,399  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/232,401  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/241,808  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/241,826  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/241,786  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/241,221  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/246,475  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/231,243  
; PRIOR FILING DATE: 2000-09-08

Query Match 2.0%; Score 17; DB 9; Length 296;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 AAATTCTAGTAAGAC 30

Db 142 AAATCTAGTAGAGC 158  
|||||  
RESULT 39  
US-10-091-572-580  
; Sequence 580 Application US/10091572  
; Publication No. US20030054373A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: P118C1  
; CURRENT APPLICATION NUMBER: US/10/091,572  
; CURRENT FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 09/764,850  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/214,886  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 60/217,487  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,758  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,963  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/217,496  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,447  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/218,290  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/225,757  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/226,868  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/216,647  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: 60/225,267  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/216,880  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: 60/225,270  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/251,869  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/235,834  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/234,274  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: 60/234,223  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: 60/228,924  
; PRIOR FILING DATE: 2000-08-30  
; PRIOR APPLICATION NUMBER: 60/224,518  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/236,369  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/224,519  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,964  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/241,809  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/249,299  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/236,327  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/241,785  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/244,617  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR FILING DATE: 2000-11-01  
; PRIOR APPLICATION NUMBER: 60/225,268  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/236,368  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/251,856  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/251,868  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/229,344  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/234,997  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: 60/229,343  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,345  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,287  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,513  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: 60/231,413  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/229,509  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: 60/236,367  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/237,039  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,038  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/236,370  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/236,802  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,037  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,040  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/240,960  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/239,935  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: 60/239,937  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: 60/241,787  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/246,474  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/246,532  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/249,216  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,210  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/226,681  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/225,759  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/225,213  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/227,182  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/225,214  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/235,836  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/230,438  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/215,135  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: 60/225,266  
; PRIOR FILING DATE: 2000-08-14



;; PRIOR APPLICATION NUMBER: 60/249,218  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,208  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,213  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,212  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,207  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,245  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,244  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,217  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,211  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,215  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,264  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,214  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,297  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/232,400  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/231,242  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/232,081  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/232,080  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/231,414  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/231,244  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/233,064  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/233,063  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/232,397  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/232,399  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/232,401  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/241,808  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/241,826  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/241,786  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/241,221  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/246,475  
;; PRIOR FILING DATE: 2000-11-08  
;; PRIOR APPLICATION NUMBER: 60/231,243  
;; PRIOR FILING DATE: 2000-09-08

Query Match 2.0%; Score 17; DB 9; Length 296;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 AAATTCCTAGTAAGAAGC 30  
| | | | | | | | | | | | | | | |  
Db 142 AAATTCCTAGTAAGAAGC 158

RESULT 40  
US-09-764-891-6723  
; Sequence 6723, Application US/09764891

; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6723  
; LENGTH: 296  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-891-6723

Query Match 2.0%; Score 17; DB 9; Length 296;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 AAATTCCTAGTAAGAAGC 30  
| | | | | | | | | | | | | | | |  
Db 142 AAATTCCTAGTAAGAAGC 158

Search completed: July 4, 2003, 03:14:33  
Job time : 161 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 01:31:27 ; Search time 1081 Seconds  
(without alignments)  
12584.837 Million cell updates/sec

Title: US-10-062-624-41  
Perfect score: 840  
Sequence: 1 atgaattataagaataattct.....ttggaatgaggttcaccttc 840

Scoring table: Oligo NUC  
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

EST:

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pln:\*
- 21: em\_gss\_vrt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_mam:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	23	2.7	619	13	BI419031
C 2	21	2.5	580	17	DR9K8T
C 3	21	2.5	624	10	AV730491
4	20	2.4	450	12	BE981973
5	20	2.4	701	10	BE577307
C 6	20	2.4	964	17	AZ539275

C	80	18	2.1	298	10	BB247871	BB247871	BB247871	BB247871	153	18	2.1	621	17	AQ484737	AQ484737	RPCI-11-2
C	81	18	2.1	301	9	AV045060	AV045060	AV045060	AV045060	154	18	2.1	622	9	AJ497575	AJ497575	AJ497575
C	82	18	2.1	311	14	BM735410	BM735410	BM735410	BM735410	155	18	2.1	631	12	BG157477	BG157477	BG157477
C	83	18	2.1	329	17	BH182650	BH182650	BH182650	BH182650	156	18	2.1	631	12	BG10982	BG10982	BG10982
C	84	18	2.1	329	17	CNS07N2X	CNS07N2X	CNS07N2X	CNS07N2X	157	18	2.1	631	12	BG936482	BG936482	BG936482
C	85	18	2.1	355	9	AI692161	AI692161	AI692161	AI692161	158	18	2.1	636	17	AZ271103	AZ271103	AZ271103
C	86	18	2.1	364	13	BM568955	BM568955	BM568955	BM568955	159	18	2.1	638	17	AZ987616	AZ987616	AZ987616
C	87	18	2.1	392	14	BM677115	BM677115	BM677115	BM677115	160	18	2.1	639	17	AQ390007	AQ390007	AQ390007
C	88	18	2.1	392	14	BM721471	BM721471	BM721471	BM721471	161	18	2.1	640	17	BU004195	BU004195	BU004195
C	89	18	2.1	395	9	AL385472	AL385472	AL385472	AL385472	162	18	2.1	646	17	BU051530	BU051530	BU051530
C	90	18	2.1	406	17	AZ033880	AZ033880	AZ033880	AZ033880	163	18	2.1	647	10	AV440067	AV440067	AV440067
C	91	18	2.1	407	12	BF802719	BF802719	BF802719	BF802719	164	18	2.1	651	10	BV616048	BV616048	BV616048
C	92	18	2.1	407	12	BE749377	BE749377	BE749377	BE749377	165	18	2.1	652	17	AZ012028	AZ012028	AZ012028
C	93	18	2.1	413	10	AW891777	AW891777	AW891777	AW891777	166	18	2.1	652	17	BE255629	BE255629	BE255629
C	94	18	2.1	418	17	AQ199334	AQ199334	AQ199334	AQ199334	167	18	2.1	660	10	BE658933	BE658933	BE658933
C	95	18	2.1	436	9	AI724677	AI724677	AI724677	AI724677	168	18	2.1	661	9	AL636140	AL636140	AL636140
C	96	18	2.1	440	17	CNS007BL	CNS007BL	CNS007BL	CNS007BL	169	18	2.1	661	10	AV821678	AV821678	AV821678
C	97	18	2.1	441	12	BE806727	BE806727	BE806727	BE806727	170	18	2.1	662	14	BQ870633	BQ870633	BQ870633
C	98	18	2.1	442	12	BE806728	BE806728	BE806728	BE806728	171	18	2.1	668	17	AZ334127	AZ334127	AZ334127
C	99	18	2.1	444	9	AU226674	AU226674	AU226674	AU226674	172	18	2.1	678	14	BU006116	BU006116	BU006116
C	100	18	2.1	448	13	BI033305	BI033305	BI033305	BI033305	173	18	2.1	679	17	BS155879	BS155879	BS155879
C	101	18	2.1	452	9	AI939345	AI939345	AI939345	AI939345	174	18	2.1	680	10	BB617739	BB617739	BB617739
C	102	18	2.1	453	17	AQ033532	AQ033532	AQ033532	AQ033532	175	18	2.1	686	13	BM163703	BM163703	BM163703
C	103	18	2.1	459	10	AV827826	AV827826	AV827826	AV827826	176	18	2.1	691	14	BU010251	BU010251	BU010251
C	104	18	2.1	462	17	AQ828865	AQ828865	AQ828865									

226	17	2.0	124	9	AU009291	AU009291	AU009291	C 299	17	2.0	339	10	BE534039
227	17	2.0	160	9	AV157539	AV157539	AV157539	C 300	17	2.0	342	9	AA943771
228	17	2.0	164	17	B84537	B84537	RPC111-26B3	C 301	17	2.0	344	12	BF537839
229	17	2.0	166	12	BG408690	BG408690	dac14608.	C 302	17	2.0	345	10	BE690597
230	17	2.0	172	17	AZ378985	AZ378985	1M0134C01	C 303	17	2.0	345	14	BQ141069
231	17	2.0	183	17	B99621	B99621	CIT-HSP-228	C 304	17	2.0	346	9	AA454406
232	17	2.0	184	10	BB056597	BB056597	BB056597	C 305	17	2.0	348	12	BF368097
233	17	2.0	192	17	BH818075	BH818075	BACFP10-D	C 306	17	2.0	348	14	H56226
234	17	2.0	198	17	AQ262934	AQ262934	CITBI-E1	C 307	17	2.0	351	17	AQ129551
235	17	2.0	205	13	BI783522	BI783522	klh33d08.Y	C 308	17	2.0	352	17	AQ710634
236	17	2.0	208	12	BF963343	BF963343	QV2-NN004	C 309	17	2.0	353	10	AV826855
237	17	2.0	216	10	AV363273	AV363273	AV363273	C 310	17	2.0	354	14	C22245
238	17	2.0	224	10	BB280367	BB280367	BB280367	C 311	17	2.0	356	17	B83164
239	17	2.0	232	12	BF451081	BF451081	u275f06.X	C 312	17	2.0	357	12	BF148479
240	17	2.0	235	10	BB012723	BB012723	BB012723	C 313	17	2.0	357	14	C22227
241	17	2.0	241	13	BM176076	BM176076	TGESTzyb2	C 314	17	2.0	360	12	BE852504
242	17	2.0	246	9	A1280294	A1280294	qm58c07.X	C 315	17	2.0	360	14	C63323
243	17	2.0	247	17	BH328600	BH328600	CH230-188	C 316	17	2.0	362	14	R09947
244	17	2.0	248	17	AZ930325	AZ930325	474.dhz52	C 317	17	2.0	364	13	RJ038787
245	17	2.0	249	14	H13783	H13783	YJ09D10.S1	C 318	17	2.0	364	13	AZ041127
246	17	2.0	252	14	BQ914709	BQ914709	QBH11119.	C 319	17	2.0	366	17	AZ474636
247	17	2.0	255	9	A1228086	A1228086	EST224781	C 320	17	2.0	367	10	AW432191
248	17	2.0	257	10	AV366676	AV366676	AV366676	C 321	17	2.0	368	10	AV530745
249	17	2.0	263	10	BB001841	BB001841	BB001841	C 322	17	2.0	368	10	AW556187
250	17	2.0	264	10	AV328074	AV328074	AV328074	C 323	17	2.0	371	10	BB792933
251	17	2.0	271	9	AV079416	AV079416	AV079416	C 324	17	2.0	372	12	BG626329
252	17	2.0	272	10	BB525869	BB525869	BB525869	C 325	17	2.0	372	17	BH189482
253	17	2.0	273	12	BF887253	BF887253	CM4-TN014	C 326	17	2.0	372	17	CNS07T9E
254	17	2.0	276	9	AV079041	AV079041	AV079041	C 327	17	2.0	374	9	AA208940
255	17	2.0	278	17	AZ649909	AZ649909	IM0519113	C 328	17	2.0	375	10	BB739531
256	17	2.0	280	9	AV023695	AV023695	AV023695	C 329	17	2.0	375	10	BB628377
257	17	2.0	280	12	BF660584	BF660584	maa1e01.	C 330	17	2.0	376	10	BB337532
258	17	2.0	282	9	AV222361	AV222361	AV222361	C 331	17	2.0	377	13	BI187600
259	17	2.0	284	9	AA589639	AA589639	v149h06.S	C 332	17	2.0	378	14	BQ235686
260	17	2.0	285	10	BB155807	BB155807	BB155807	C 333	17	2.0	378	17	AQ087936
261	17	2.0	287	10	BB316439	BB316439	BB316439	C 334	17	2.0	379	13	BI187238
262	17	2.0	290	9	AL655825	AL655825	AL655825	C 335	17	2.0	379	17	BH353609
263	17	2.0	291	12	BF450442	BF450442	u278g10.X	C 336	17	2.0	380	14	R77594
264	17	2.0	293	10	AV298003	AV298003	AV298003	C 337	17	2.0	382	10	BB687150
265	17	2.0	293	14	BQ106774	BQ106774	NX1V089.C	C 338	17	2.0	382	12	BF151382
266	17	2.0	294	9	A1878296	A1878296	fc59a04.Y	C 339	17	2.0	383	10	BB828859
267	17	2.0	297	9	AV050784	AV050784	AV050784	C 340	17	2.0	383	14	D61824
268	17	2.0	298	10	AV418517	AV418517	AV418517	C 341	17	2.0	383	17	AQ067549
269	17	2.0	298	10	BE634577	BE634577	uw83e01.X	C 342	17	2.0	385	12	BF147298
270	17	2.0	300	14	Z97770	Z97770	CFZ97770.Do	C 343	17	2.0	385	17	AQ188845
271	17	2.0	301	9	AL836574	AL836574	AL836574	C 344	17	2.0	386	17	AQ020988
272	17	2.0	301	9	AL836601	AL836601	AL836601	C 345	17	2.0	387	10	BB817350
273	17	2.0	301	9	AL836606	AL836606	AL836606	C 346	17	2.0	387	13	BI187599
274	17	2.0	301	10	AW092112	AW092112	EST285388	C 347	17	2.0	388	9	AA721858
275	17	2.0	305	14	BQ968309	BQ968309	QBH33105.	C 348	17	2.0	389	10	AW588184
276	17	2.0	306	9	AV096566	AV096566	AV096566	C 349	17	2.0	390	17	AQ190889
277	17	2.0	309	9	AU018275	AU018275	AU018275	C 350	17	2.0	391	9	A1089205
278	17	2.0	310	17	AQ478193	AQ478193	RPCI-11-2	C 351	17	2.0	391	17	BH348800
279	17	2.0	312	9	A1301387	A1301387	qm71C10.X	C 352	17	2.0	392	13	BI187237
280	17	2.0	312	12	BG147346	BG147346	mac08C05.	C 353	17	2.0	393	12	BB812410
281	17	2.0	315	17	AQ614234	AQ614234	RPCI93-Dp	C 354	17	2.0	394	12	BF633874
282	17	2.0	316	10	AW094460	AW094460	EST287640	C 355	17	2.0	394	13	BG951998
283	17	2.0	318	10	BE690980	BE690980	uw54b10.X	C 356	17	2.0	397	10	AW056887
284	17	2.0	318	12	BE921501	BE921501	EST425270	C 357	17	2.0	397	10	BB671768
285	17	2.0	323	9	AL168954	AL168954	mr31C01.X	C 358	17	2.0	397	17	BH858472
286	17	2.0	323	10	AW319719	AW319719	un07a07.X	C 359	17	2.0	400	17	BB685195
287	17	2.0	323	13	BM279975	BM279975	zah6155.Z	C 360	17	2.0	400	10	BB816567
288	17	2.0	324	12	BF148314	BF148314	uy79d03.X	C 361	17	2.0	400	13	BJ378158
289	17	2.0	327	9	A1016180	A1016180	ot73g11.S	C 362	17	2.0	402	12	BF822329
290	17	2.0	327	10	BB448262	BB448262	BB448262	C 363	17	2.0	404	9	AL599362
291	17	2.0	330	10	BB251232	BB251232	BB251232	C 364	17	2.0	405	14	R07850
292	17	2.0	330	17	AQ060470	AQ060470	CIT-HSP-2	C 365	17	2.0	407	10	BB735233
293	17	2.0	330	17	AZ889512	AZ889512	RPCI-24-1	C 366	17	2.0	409	17	AQ811466
294	17	2.0	333	9	AV009123	AV009123	AV009123	C 367	17	2.0	409	17	AQ149375
295	17	2.0	334	17	AQ114123	AQ114123	CIT-HSP-2	C 368	17	2.0	410	13	BI266118
296	17	2.0	335	10	BB113459	BB113459	BB113459	C 369	17	2.0	411	10	BB743376
297	17	2.0	338	10	BB453645	BB453645	BB453645	C 370	17	2.0	412	9	AA676689
298	17	2.0	338	10	BB764594	BB764594	BB764594	C 371	17	2.0	412	10	BB810565

372	17	2.0	414	9	AU229051	AU229051	AU229051	445	17	2.0	468	13	BM207236	BM207236	BM207236	CO610A05
373	17	2.0	414	17	A2914234	RPCI-24-1	A2914234	C 446	17	2.0	468	17	AZ005707	AZ005707	AZ005707	RPCI-23-3
374	17	2.0	415	9	A278623	33217 Lam	A278623	C 447	17	2.0	469	10	BE447029	BE447029	BE447029	u55611.x
375	17	2.0	417	10	AW741739	u59e05.x	AW741739	C 448	17	2.0	470	10	BE785677	BE785677	BE785677	BM785677
376	17	2.0	417	10	BE789412	BB789412	BE789412	C 449	17	2.0	470	17	AZ786205	AZ786205	2M0031N05	AZ786205
377	17	2.0	417	10	BE826127	BB826127	BE826127	C 450	17	2.0	471	17	AQ841649	AQ841649	TI137402b	AQ841649
378	17	2.0	418	10	BE819317	BB819317	BE819317	C 451	17	2.0	471	17	BG134051	BG134051	SALK_0354	BG134051
379	17	2.0	419	9	AI256152	u195f05.x	AI256152	C 452	17	2.0	473	12	BG136784	BG136784	EST477226	AI256152
380	17	2.0	419	13	BM228353	K0259H04-	BM228353	C 453	17	2.0	473	17	AZ447718	AZ447718	1M0245E06	BM228353
381	17	2.0	420	10	BE789121	BB789121	BE789121	C 454	17	2.0	474	10	BE446867	BE446867	u888603.x	BE789121
382	17	2.0	423	17	AQ254167	HS_3020_A	AQ254167	C 455	17	2.0	475	9	AI183927	AI183927	qd69a08.x	AQ254167
383	17	2.0	424	10	BE677341	BB677341	BE677341	C 456	17	2.0	475	10	BE837156	BE837156	BB837156	BE677341
384	17	2.0	424	17	AZ035620	RPCI-23-3	AZ035620	C 457	17	2.0	475	10	BE868628	BE868628	uv74803.x	AZ035620
385	17	2.0	425	17	AZ595397	1M0408E01	AZ595397	C 458	17	2.0	476	10	BE822390	BE822390	BB822390	AZ595397
386	17	2.0	426	12	BG669843	DRNAKH11	BG669843	C 459	17	2.0	478	10	BE744408	BE744408	BB744408	BG669843
387	17	2.0	427	10	AW552208	L0209G07-	AW552208	C 460	17	2.0	479	9	AA274068	AA274068	vb93g09.x	AW552208
388	17	2.0	427	10	BE780518	BB780518	BE780518	C 461	17	2.0	479	10	BE729867	BE729867	BB729867	BE780518
389	17	2.0	427	12	BE847720	uv93c01.y	BE847720	C 462	17	2.0	479	14	BQ106340	BQ106340	fc1198.e	BE847720
390	17	2.0	428	10	BE825870	BB825870	BE825870	C 463	17	2.0	479	17	CNS02EPA	CNS02EPA	retroacodon	BE825870
391	17	2.0	428	14	T99406	ye64c06.r1	T99406	C 464	17	2.0	479	17	AQ660820	AQ660820	Sheared D	T99406
392	17	2.0	429	10	BE750853	BB750853	BE750853	C 465	17	2.0	480	9	AU211297	AU211297	AU211297	BE750853
393	17	2.0	431	9	AA231400	mw31e11.x	AA231400	C 466	17	2.0	480	13	BM522286	BM522286	ESSU0824	AA231400
394	17	2.0	431	9	AA285461	vb90h12.x	AA285461	C 467	17	2.0	480	17	AZ168998	AZ168998	SP_0112_B	AA285461
395	17	2.0	433	10	BE825578	BB825578	BE825578	C 468	17	2.0	480	17	AQ0505810	AQ0505810	RFC1-11-2	BE825578
396	17	2.0	434	12	BF017339	uw74f06.x	BF017339	C 469	17	2.0	483	12	BG813997	BG813997	gaif62b07.	BF01733



664	17	2.0	596	9	AJ425151	737	17	2.0	655	17	AG162672	AG162672	Pan trogl
665	17	2.0	596	13	BM230375	C 738	17	2.0	656	9	AA265257	AA265257	mr21f12.r
666	17	2.0	597	17	AQ921177	C 739	17	2.0	657	10	AW633588	AW633588	NF067H1LS
667	17	2.0	598	13	B1372599	C 740	17	2.0	657	10	BB392582	BB392582	BB392582
668	17	2.0	598	13	BM166940	C 741	17	2.0	658	12	BB630400	BB630400	BB630400
669	17	2.0	598	17	BM575035	C 742	17	2.0	658	12	BF636445	BF636445	BF636445
670	17	2.0	600	13	B11990305	C 743	17	2.0	659	14	BM559206	BM559206	BM559206
671	17	2.0	600	17	BM210019	C 744	17	2.0	659	14	BQ157335	BQ157335	BM559206
672	17	2.0	600	17	AQ449575	C 745	17	2.0	660	13	BM243957	BM243957	BM559206
673	17	2.0	601	13	B1348551	C 746	17	2.0	660	13	BM243957	BM243957	BM559206
674	17	2.0	601	17	AQ250194	C 747	17	2.0	661	17	AZ527102	AZ527102	BM559206
675	17	2.0	602	9	AI226404	C 748	17	2.0	662	9	AU091558	AU091558	BM559206
676	17	2.0	602	9	AI226404	C 749	17	2.0	662	9	AU091558	AU091558	BM559206
677	17	2.0	602	9	AL700725	C 750	17	2.0	663	12	BF393631	BF393631	BM559206
678	17	2.0	602	17	BM398648	C 751	17	2.0	663	12	BF393631	BF393631	BM559206
679	17	2.0	605	10	BM247041	C 752	17	2.0	664	14	BQ968885	BQ968885	BM559206
680	17	2.0	605	13	BM244228	C 753	17	2.0	664	14	BQ968885	BQ968885	BM559206
681	17	2.0	605	17	BM021810	C 754	17	2.0	665	12	BF649281	BF649281	BM559206
682	17	2.0	607	17	BM874069	C 755	17	2.0	665	12	BF649281	BF649281	BM559206
683	17	2.0	608	10	AV871047	C 756	17	2.0	666	9	AU032202	AU032202	BM559206
684	17	2.0	609	17	BM276111	C 757	17	2.0	666	9	AU032202	AU032202	BM559206
685	17	2.0	609	17	AQ256311	C 758	17	2.0	667	14	BQ968885	BQ968885	BM559206
686	17	2.0	609	17	AL753893	C 759	17	2.0	668	12	BF649281	BF649281	BM559206
687	17	2.0	611	9	AU024178	C 760	17	2.0	668	12	BF649281	BF649281	BM559206
688	17	2.0	613	13	B1802007	C 761	17	2.0	669	14	BQ968885	BQ968885	BM559206
689	17	2.0	613	14	BQ177006	C 762	17	2.0	670	13	BM157335	BM157335	BM559206
690	17	2.0	613	17	A2454207	C 763	17	2.0	670	13	BM157335	BM157335	BM559206
691	17	2.0	614	13	BM594186	C 764	17	2.0	671	13	BM164632	BM164632	BM559206
692	17	2.0	615	17	AQ415984	C 765	17	2.0	671	13	BM164632	BM164632	BM559206
693	17	2.0	616	13	BQ378087	C 766	17	2.0	672	13	BM164632	BM164632	BM559206
694	17	2.0	617	17	A2974077	C 767	17	2.0	672	13	BM164632	BM164632	BM559206
695	17	2.0	619	17	FR0022729	C 768	17	2.0	673	10	AW555066	AW555066	BM559206
696	17	2.0	621	9	AL637714	C 769	17	2.0	674	17	AZ405167	AZ405167	BM559206
697	17	2.0	622	9	AA184209	C 770	17	2.0	674	17	AZ405167	AZ405167	BM559206
698	17	2.0	622	10	BE324946	C 771	17	2.0	675	9	AA009119	AA009119	BM559206
699	17	2.0	626	10	BQ205932	C 772	17	2.0	675	12	BG525471	BG525471	BM559206
700	17	2.0	626	17	AG025932	C 773	17	2.0	676	12	BG525471	BG525471	BM559206
701	17	2.0	624	9	AL673173	C 774	17	2.0	677	14	BQ205932	BQ205932	BM559206
702	17	2.0	624	17	BH112446	C 775	17	2.0	677	14	BQ205932	BQ205932	BM559206
703	17	2.0	625	9	AA638517	C 776	17	2.0	678	14	BQ205932	BQ205932	BM559206
704	17	2.0	625	13	BM225019	C 777	17	2.0	679	13	BQ205932	BQ205932	BM559206
705	17	2.0	626	10	BQ203429	C 778	17	2.0	679	13	BQ205932	BQ205932	BM559206
706	17	2.0	626	14	BQ875589	C 779	17	2.0	680	14	BQ967420	BQ967420	BM559206
707	17	2.0	627	13	BM228308	C 780	17	2.0	680	14	BQ967420	BQ967420	BM559206
708	17	2.0	628	13	B1150184	C 781	17	2.0	681	14	BQ967420	BQ967420	BM559206
709	17	2.0	629	10	AV846144	C 782	17	2.0	681	14	BQ967420	BQ967420	BM559206
710	17	2.0	630	17	A2402829	C 783	17	2.0	682	14	BQ967420	BQ967420	BM559206
711	17	2.0	630	17	AQ482224	C 784	17	2.0	683	12	BG191161	BG191161	BM559206
712	17	2.0	632	13	BM604669	C 785	17	2.0	683	12	BG191161	BG191161	BM559206
713	17	2.0	632	17	AZ015086	C 786	17	2.0	684	14	BQ967420	BQ967420	BM559206
714	17	2.0	632	17	AZ187134	C 787	17	2.0	685	14	BQ967420	BQ967420	BM559206
715	17	2.0	636	13	BM159685	C 788	17	2.0	685	14	BQ967420	BQ967420	BM559206
716	17	2.0	637	14	BQ149511	C 789	17	2.0	686	14	BQ967420	BQ967420	BM559206
717	17	2.0	638	12	BF121233	C 790	17	2.0	687	14	BQ967420	BQ967420	BM559206
718	17	2.0	639	17	AZ961160	C 791	17	2.0	688	14	BQ967420	BQ967420	BM559206
719	17	2.0	640	17	AZ968743	C 792	17	2.0	689	14	BQ967420	BQ967420	BM559206
720	17	2.0	641	12	BF646730	C 793	17	2.0	690	14	BQ967420	BQ967420	BM559206
721	17	2.0	642	10	BM659286	C 794	17	2.0	691	14	BQ967420	BQ967420	BM559206
722	17	2.0	642	14	BQ037117	C 795	17	2.0	692	14	BQ967420	BQ967420	BM559206
723	17	2.0	642	14	BQ414561	C 796	17	2.0	693	12	BQ967420	BQ967420	BM559206
724	17	2.0	642	17	AZ602407	C 797	17	2.0	694	14	BQ967420	BQ967420	BM559206
725	17	2.0	643	10	AV204473	C 798	17	2.0	695	14	BQ967420	BQ967420	BM559206
726	17	2.0	643	10	BM657541	C 799	17	2.0	696	14	BQ967420	BQ967420	BM559206
727	17	2.0	647	10	BM204425	C 800	17	2.0	697	14	BQ967420	BQ967420	BM559206
728	17	2.0	648	9	AA720375	C 801	17	2.0	698	14	BQ967420	BQ967420	BM559206
729	17	2.0	649	9	AI405703	C 802	17	2.0	699	14	BQ967420	BQ967420	BM559206
730	17	2.0	649	13	BQ081791	C 803	17	2.0	700	14	BQ967420	BQ967420	BM559206
731	17	2.0	650	13	BQ449762	C 804	17	2.0	701	14	BQ967420	BQ967420	BM559206
732	17	2.0	653	13	BM968329	C 805	17	2.0	702	17	BH522286	BH522286	BM559206
733	17	2.0	653	14	BM968329	C 806	17	2.0	702	17	BH522286	BH522286	BM559206
734	17	2.0	654	10	BM038325	C 807	17	2.0	702	17	BH522286	BH522286	BM559206
735	17	2.0	654	10	BM038325	C 808	17	2.0	702	17	BH522286	BH522286	BM559206
736	17	2.0	654	17	AZ506841	C 809	17	2.0	702	17	BH522286	BH522286	BM559206



C 810	17	703	17	BH399878	AG-ND-162	C 883	17	2.0	816	17	AQ249532	T17A11-T7
811	17	705	17	BH739126	BHRL32TR	C 884	17	2.0	820	17	BH534577	BH534577 BOCF93TR
812	17	706	14	BQ195094	UI-R-CN1-	C 885	17	2.0	822	14	BQ165667	BQ165667 BOCF611536
813	17	707	13	BI305817	BI305817 NL_1 N16	C 886	17	2.0	825	12	BG197197	BG197197 RST16436
814	17	707	17	A22113038	SheaTed D	C 887	17	2.0	826	12	BG595797	BG595797 RST494475
815	17	708	12	BG5313358	BG5313358 602559605	C 888	17	2.0	826	17	BH459924	BH459924 BOMJA77TF
816	17	709	12	BG616033	BG616033 602643478	C 889	17	2.0	827	13	BI821861	BI821861 603035751
817	17	709	12	BG646338	BG646338 EST507957	C 890	17	2.0	831	12	BG068761	BG068761 H3069A08-
818	17	709	17	AZ603624	IM0423M02	C 891	17	2.0	831	17	AQ751815	AQ751815 HS 5569 B
819	17	710	17	CNS028U2	AL190307 Tetraodon	C 892	17	2.0	831	17	AZ900733	RPCI-24-1
C 820	17	711	9	AUI65386	AUI65386	C 893	17	2.0	832	17	AZ527717	ENTDA55TF
821	17	711	14	BQ915968	BQ915968 OHL61813	C 894	17	2.0	839	17	AZ688121	ENTLN94TF
822	17	720	9	AL633809	AL633809	C 895	17	2.0	839	17	BI332387	BG3983580
C 823	17	721	9	AU094060	AU094060 AU094060	C 896	17	2.0	841	12	BF241984	BF241984 601878320
C 824	17	722	17	AZ915157	RPCI-24-1	C 897	17	2.0	842	12	BG195167	BG195167 RST14346
C 825	17	725	10	BE284216	601099146	C 898	17	2.0	848	17	CNS02QV1	Tetraodon
C 826	17	725	12	BF183567	BF183567 601809884	C 899	17	2.0	850	17	BH455721	BOGGP36TR
C 827	17	727	17	AQ842751	CpG1265A	C 900	17	2.0	855	17	AZ544806	ENTFX32TF
C 828	17	727	17	AQ879104	AQ879104 HS_3184 A	C 901	17	2.0	855	17	CNS01Q3T	AL155625 Anopheles
C 829	17	728	12	BG632943	BG632943 GHI6652.3	C 902	17	2.0	856	17	AQ748318	HS 5535 A
830	17	729	14	BQ966688	BQ966688 QHB27J11-	C 903	17	2.0	861	12	BG426937	602493125
C 831	17	731	12	BG297775	BG297775 602393824	C 904	17	2.0	861	12	BE868989	601444780
C 832	17	733	13	BJ375849	BJ375849	C 905	17	2.0	865	17	AQ745624	HS_2273 A
C 833	17	733	17	BH545373	BH545373 BQGLT32TR	C 906	17	2.0	867	17	CNS04JHM	Tetraodon
C 834	17	734	17	BH565509	BH565509 BOMNU36TF	C 907	17	2.0	871	17	AZ670777	ENTHR65TR
C 835	17	737	17	AQ888972	AQ888972 HS_3100 A	C 908	17	2.0	875	17	BH159925	ENTTR78TF
C 836	17	739	14	BQ444070	BQ444070 UI-M-EX0-	C 909	17	2.0	875	17	BH654077	BOMGT70TF
C 837	17	744	17	AQ842833	CpG1265B	C 910	17	2.0	877	13	BI665667	603289663
C 838	17	745	17	CNS03D0H	AL238490 Tetraodon	C 911	17	2.0	878	17	AZ682831	ENTKG32TR
C 839	17	751	17	BH244791	BH244791 AUIJA58TF	C 912	17	2.0	878	17	BH130554	ENTQO72TR
C 840	17	753	14	BQ969770	BQ969770 QHB39H22-	C 913	17	2.0	883	13	BI108530	ENTSU09TR
C 841	17	754	14	BM680271	UI-E-E01-	C 914	17	2.0	883	13	BI158341	ENTSU09TR
C 842	17	754	17	BH376762	BH376762 AG-ND-172	C 915	17	2.0	884	12	BG573758	602594842
C 843	17	759	12	BF131704	BF131704 601820674	C 916	17	2.0	885	13	BI853134	603379673
844	17	759	14	BQ165498	BQ165498 ESP611367	C 917	17	2.0	886	12	BG415096	BG415096 HVSMEX000
845	17	759	17	AZ191535	AZ191535 SP_1019 B	C 918	17	2.0	886	12	BG415096	BG415096 HVSMEX000
846	17	760	17	AQ743302	AQ743302 HS_5387 B	C 919	17	2.0	887	12	BF608390	MY1_00133
C 847	17	760	17	BH046920	BE564211 601343005	C 920	17	2.0	887	17	AZ549067	ENTDG19TR
848	17	764	10	BE564211	601343005	C 921	17	2.0	889	17	AZ532105	ENTB74TR
849	17	764	12	BF132063	BF132063 601821067	C 922	17	2.0	890	17	CNS02Z01	Tetraodon
C 850	17	764	12	BF139145	BF139145 601784185	C 923	17	2.0	892	17	AZ534663	ENTCW55TF
C 851	17	766	17	AQ961079	AQ961079 LERFU06TF	C 924	17	2.0	892	17	BH353495	CH230-55B
C 852	17	767	9	AL667815	AL667815 AL667815	C 925	17	2.0	893	17	AQ872507	nba60050F
853	17	767	10	BE130387	L48-561T3	C 926	17	2.0	894	17	AZ548062	ENTET45TF
854	17	768	13	BM164404	BM164404 EST656927	C 927	17	2.0	894	17	BH132534	ENTNQ46TR
855	17	768	14	BQ968829	BQ968829 QHB35E18-	C 928	17	2.0	895	17	AZ550279	ENTCZ47TR
856	17	773	17	BH588734	BH588734 BQGT69TF	C 929	17	2.0	895	17	AZ688731	ENTZ60TF
C 857	17	775	14	BQ137146	BQ137146 NF067H11S	C 930	17	2.0	897	12	BF142967	BF142967 601791746
C 858	17	778	17	AQ745800	AQ745800 HS_2273 A	C 931	17	2.0	900	13	BI853584	603380706
C 859	17	778	17	BH143285	BH143285 TDGDY36TH	C 932	17	2.0	900	17	BH150137	ENTPU45TR
C 860	17	779	17	BH342662	BH342662 CH230-183	C 933	17	2.0	902	17	AZ676768	ENTLU11TF
861	17	779	17	CNS0700N	AL423885 T3 end of	C 934	17	2.0	903	14	BQ936806	BQ936806 ACENCCOURT
862	17	780	13	BM171040	BM171040 EST573563	C 935	17	2.0	904	17	BH133978	ACENCCOURT
863	17	780	17	BH794812	BH794812 ME_MBA000	C 936	17	2.0	905	12	BG529186	602558256
C 864	17	782	12	BG220112	BG220112 RST39884	C 937	17	2.0	906	17	BH152565	ENTPW30TR
C 865	17	783	14	BQ655600	BQ655600 NKRVO96 F	C 938	17	2.0	908	17	BH146972	ENTPQ32TR
C 866	17	784	17	BH578454	BH578454 BOMPR49TR	C 939	17	2.0	909	17	BH134223	ENTNX68TR
C 867	17	789	12	BF218716	BF218716 601884942	C 940	17	2.0	910	14	BQ960247	ACENCCOURT
868	17	789	17	CNS01MRB	AL110768 Tetraodon	C 941	17	2.0	910	17	CNS01HXV	AL115028 Anopheles
869	17	791	12	BG202888	BG202888 RST22256	C 942	17	2.0	911	17	BH165898	ENTSH20TR
C 870	17	791	17	CNS03WGU	AL263703 Tetraodon	C 943	17	2.0	911	17	CNS03L48	AL248993 Tetraodon
C 871	17	792	13	BI186279	BI186279 UNL-P-FN-	C 944	17	2.0	912	13	BI648340	603278291
C 872	17	794	13	BI687489	BI687489 603315042	C 945	17	2.0	916	17	CNS01N9Y	AL151959 Anopheles
873	17	799	17	BH377040	BH377040 AG-ND-171	C 946	17	2.0	918	12	BF570868	BF570868 602076044
C 874	17	801	13	BI308329	BI308329 EST529739	C 947	17	2.0	919	14	BQ864323	BQ864323 ACENCCOURT
C 875	17	801	17	BH661909	BH661909 BOMNH96TF	C 948	17	2.0	920	17	BH147810	ENTPN23TR
C 876	17	805	17	BH244772	BH244772 AUIJA10TR	C 949	17	2.0	922	13	BM357973	GA_Ea000
C 877	17	805	17	BH586913	BH586913 BOHAG22TR	C 950	17	2.0	925	17	AZ528020	ENTBG48TF
878	17	806	12	BG620889	BG620889 602617226	C 951	17	2.0	926	13	BI524333	603052187
879	17	806	13	BI255914	BI255914 602976536	C 952	17	2.0	926	17	BH149422	ENTQO2TR
C 880	17	809	12	BG518514	BG518514 602578246	C 953	17	2.0	932	12	BG184614	RST3676 A
C 881	17	810	17	AQ855294	CGI825B	C 954	17	2.0	934	12	BG741310	602634445
C 882	17	814	17	BH086558	BH086558 RPCI-24-2	C 955	17	2.0	936	17	BH146771	ENTPR28TR

```

956 17 2.0 937 12 BF536470
c 957 17 2.0 938 17 BHI60048
c 958 17 2.0 939 12 BE789702
c 959 17 2.0 943 17 BHI66861
960 17 2.0 948 14 W33601
961 17 2.0 949 17 CNS01PRV
c 962 17 2.0 952 12 BF782584
963 17 2.0 954 12 BG671449
c 964 17 2.0 954 12 BF301373
965 17 2.0 954 17 CNS01TGA
c 966 17 2.0 956 17 AQ899342
c 967 17 2.0 965 17 CNS02XJL
c 968 17 2.0 968 12 BF585159
969 17 2.0 975 12 BF159660
970 17 2.0 980 14 BQ220935
971 17 2.0 982 17 CNS05PCW
972 17 2.0 984 13 BM399836
973 17 2.0 991 14 BQ073031
974 17 2.0 994 14 BQ896242
975 17 2.0 997 14 BM904821
976 17 2.0 1005 11 AK017662
c 977 17 2.0 1005 17 CNS073VI
978 17 2.0 1014 14 BQ963228
c 979 17 2.0 1021 17 BHI52261
980 17 2.0 1027 17 CNS077AG
981 17 2.0 1028 17 CNS059RO
982 17 2.0 1050 14 BQ922341
983 17 2.0 1052 12 BF166343
c 984 17 2.0 1098 13 BM547933
c 985 17 2.0 1101 17 CNS04MX6
c 986 17 2.0 1117 12 BG254401
987 17 2.0 1118 12 BG426867
c 988 17 2.0 1131 10 BS614536
c 989 17 2.0 1165 17 AG108467
c 990 17 2.0 1266 13 BM461290
991 17 2.0 1329 11 AK015926
992 17 2.0 1334 10 BS543479
c 993 17 2.0 1349 13 BM546741
c 994 17 2.0 1398 17 P7019123R
995 17 2.0 1437 12 BE965332
996 17 2.0 1582 11 AK007706
c 997 17 2.0 1608 10 BE618396
c 998 17 2.0 2278 11 AK002792
999 17 2.0 2478 11 BC006043
c1000 17 2.0 2615 11 AK016581

```

## ALIGNMENTS

```

RESULT 1
BI419031/c 619 bp mRNA linear EST 15-AUG-2001
LOCUS LjNEST26f3r Lotus japonicus nodule library 5 and 7 week-old Lotus
DEFINITION japonicus cDNA 5', mRNA sequence.
ACCESSION BI419031
VERSION BI419031.1 GI:15190054
KEYWORDS EST.
SOURCE Lotus japonicus.
ORGANISM Lotus japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
Lotus.
1 (bases 1 to 619)
Colebatch, G., Freund, S., Trevaskis, B and Udvardi, M.
Lotus japonicus root nodule ESTs: tools for functional genomics
Unpublished (2000).
Contact: Udvardi MK
Molecular Plant Nutrition
Max Planck Institute of Molecular Plant Physiology
Am Muehlenberg 1, 14476 Golm, Germany
Fax: 49 331 567 8250

```

```

Email: udvardi@pimp-golm.mpg.de
Seq primer: T7
High quality sequence stop: 619.
Location/Qualifiers
FEATURES
source
1..619
/organism="Lotus japonicus"
/cultivar="Gifu (B-129)"
/db_xref="taxon:34305"
/clone_lib="Lotus japonicus nodule library 5 and 7
week-old"
/dev stage="5 and 7 week-old plants"
/notes="Organ: Nodule; Vector: pSPORT1; Site 1: Salt;
Site 2: NotI; The library was prepared using mRNA
extracted from nodules of 5 and 7 week-old Lotus plants.
Nodules were induced by, and contained Mesorhizobium
strain R7A."
BASE COUNT 175 a 137 c 133 g 174 t
ORIGIN
Query Match 2.7%; Score 23; DB 13; Length 619;
Best Local Similarity 100.0%; Pred. No. 0.57;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 216 TTTCGGACTAAAGAAAGATGGTG 238
Db 483 TTTCGGACTAAAGAAAGATGGTG 461
RESULT 2
DR9K8T 580 bp DNA linear GSS 06-JUN-2002
LOCUS Danio rerio genomic clone DKey-9K8, genomic survey sequence.
DEFINITION Danio rerio genomic clone DKey-9K8, genomic survey sequence.
ACCESSION AL737825
VERSION AL737825.1 GI:21352575
KEYWORDS GSS.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 580)
AUTHORS Humphray, S.J., Huckle, E. and Hunt, S.E.
TITLE Direct Submission
JOURNAL Submitted (06-JUN-2002) The Sanger Institute, Wellcome Trust Genome
Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
humquerry@sanger.ac.uk Unpublished
This sequence was generated from the T7 end of BAC 9K8. 9K8 is part
of the Daniokey Pilot BAC Library created by R. Plasterk and N.V.
Keygene.
Further details: http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES
source
1..580
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="DKey-9K8"
/tissue_type="Testis"
/notes="Vector: pindigoBAC-536"
BASE COUNT 188 a 65 c 100 g 226 t
ORIGIN
Query Match 2.5%; Score 21; DB 17; Length 580;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 284 ATTTTCAAAATAACTTAATAT 304
Db 196 ATTTTCAAAATAACTTAATAT 176
RESULT 3
AV730491 624 bp mRNA linear EST 17-OCT-2000
LOCUS AV730491
DEFINITION AV730491 HTF Homo sapiens cDNA clone HTFAAB02 5', mRNA sequence.

```

```

ACCESSION AV730491
VERSION AV730491.1 GI:10839912
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 624)
Gu.Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
Li.N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
Chen,J., Chen,Z. and Han,Z.
Homo sapiens cDNA HTP clones
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhongjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
Location/Qualifiers
1..624
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTFAAB02"
/clone_lib="HTF"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 235 a 102 c 107 g 180 t
ORIGIN
Query Match 2.5%; Score 21; DB 10; Length 624;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 419 AATACTATAACATTTCAT 439
|||||
Db 442 AATACTATAACATTTCAT 462

RESULT 4
BE991973
LOCUS BE991973
DEFINITION UI-M-B21-beb-e-12-0-UI.s1 NIH_BMAP_MHI2_S1 Mus musculus cDNA clone
EST.
ACCESSION BE991973
VERSION BE991973
KEYWORDS BE991973.1 GI:10672355
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 450)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
hippocampus tissue cDNA Library Preparation: M.B. Soares Lab Clone
distribution: Researchers may obtain BMAP cDNA clones from RESEARCH
GENETICS. It should be noted that Bento Soares is generating a
small number of additional specialized non-redundant arrays of BMAP
cDNAs whose availability will be considered under appropriate and
limited collaborative arrangements
Seq primer: M13 Forward
POLYA=Yes.
FEATURES
Location/Qualifiers
1..450
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="UI-M-B21-beb-e-12-0-UI"
/clone_lib="NIH_BMAP_MHI2_S1"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The
NIH BMAP_MHI2_S1 library is a subtracted library derived
from NIH_BMAP_MHI2. NIH_BMAP_MHI2 is a library derived
from mouse hippocampus tissue. For a detailed description
of the library from which this clone was derived, please
visit our web site at brainest.eng.uiowa.edu.
TAG_LIB=NIH_BMAP_MHI2_S1
TAG_TISSUE=hippocampus
TAG_SEQ=TAGTC"
BASE COUNT 174 a 121 c 87 g 68 t
ORIGIN
Query Match 2.4%; Score 20; DB 12; Length 450;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 CAACAATTTATCCAAAAA 389
|||||
Db 246 CAACAATTTATCCAAAAA 265

RESULT 5
BE577307
LOCUS BE577307
DEFINITION BE577307.1 GI:9827106
EST.
ACCESSION BE577307
VERSION BE577307.1
KEYWORDS BE577307.1
SOURCE common iceplant.
ORGANISM Mesembryanthemum crystallinum
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
1 (bases 1 to 701) Caryophyllidae; Caryophyllales; Alzooaceae; Mesembryanthemum.
Cushman, J.C.
TITLE An expressed sequence tag database for the common ice plant,
Mesembryanthemum crystallinum
JOURNAL Unpublished (1997)
COMMENT Contact: Cushman JC
Department of Biochemistry
University of Nevada
MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcushman@unr.edu
PCR Primers
FORWARD: T7
BACKWARD: T3
Plate: L48-22 row: C column: 9
Seq primer: T3
High quality sequence stop: 350

```

POLYA=NO. Location/Qualifiers  
 1. .701  
 /organism="Mesembryanthemum crystallinum"  
 /db\_xref="taxon:3544"  
 /clone="L48-2133"  
 /clone\_lib="Ice plant Lambda Uni-Zap XR expression library"  
 /48 hours NaCl treatment"  
 /tissue\_type="Leaf, 48 h 0.4M NaCl"  
 /dev\_stage="Six week old"  
 /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site\_1:  
 EcoRI; Site\_2: XhoI"  
 BASE COUNT 199 a 151 c 221 g 130 t  
 ORIGIN

Query Match 2.4%; Score 20; DB 10; Length 701;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 356 TTGAAGCTCATATCAACAA 375  
 Db 12 TTGAAGCTCATATCAACAA 31

RESULT 6  
 AZ539275/c  
 LOCUS  
 DEFINITION ENTM15TF Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.  
 ACCESSION AZ539275  
 VERSION AZ539275.1 GI:11144997  
 KEYWORDS GSS.  
 SOURCE Entamoeba histolytica.  
 ORGANISM Entamoeba histolytica  
 Eukaryota; Entamoebidae; Entamoeba.  
 REFERENCE 1 (bases 1 to 964)  
 AUTHORS Loftus B., Van Aken S. and Fraser C.  
 TITLE Determination of clone end sequences from Entamoeba histolytica  
 JOURNAL HMI:IMSS sheared DNA library  
 COMMENT Unpublished (2000)  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: b.loftus@tigr.org  
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library  
 Seq primer: M13-Forward  
 Class: shotgun  
 High quality sequence start: 157  
 High quality sequence stop: 383.

FEATURES  
 source  
 1. .964  
 /organism="Entamoeba histolytica"  
 /strain="HMI:IMSS"  
 /db\_xref="taxon:5759"  
 /clone\_lib="Entamoeba histolytica Sheared DNA"  
 /note="Vector: pHO51; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrel, Oxford University Press, 1999)."  
 BASE COUNT 370 a 97 c 187 g 310 t

## ORIGIN

Query Match 2.4%; Score 20; DB 17; Length 964;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 87 TTCAGAACTAATGATAACA 106  
 Db 170 TTCAGAACTAATGATAACA 151

## RESULT 7

AU074672  
 LOCUS  
 DEFINITION AU074672 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium discoideum cDNA clone SSL493, mRNA sequence.

ACCESSION AU074672  
 VERSION AU074672.1 GI:5181093  
 KEYWORDS EST.  
 SOURCE Dictyostelium discoideum.  
 ORGANISM Dictyostelium discoideum  
 Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
 REFERENCE 1 (bases 1 to 240)  
 AUTHORS Urushihara, H.  
 TITLE Developmental cDNA in Dictyostelium discoideum (1999)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Hideko Urushihara  
 Institute of Biological Sciences  
 University of Tsukuba  
 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan  
 Tel: 81-298-53-4664  
 Fax: 81-298-53-6614  
 Email: hideko@biol.tsukuba.ac.jp  
 PROJECT = Dictyostelium discoideum cDNA project in Japan.

## FEATURES

Location/Qualifiers  
 source  
 1. .240  
 /organism="Dictyostelium discoideum"  
 /strain="AX4"  
 /db\_xref="taxon:44689"  
 /clone="SSL493"  
 /clone\_lib="Dictyostelium discoideum SS (H.Urushihara)"  
 /dev\_stage="slug"  
 BASE COUNT 99 a 46 c 40 g 55 t  
 ORIGIN

Query Match 2.3%; Score 19; DB 9; Length 240;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 401 ATGATCTGTAATGGTGA 419

Db 62 ATGATCTGTAATGGTGA 80

## RESULT 8

AW288561  
 LOCUS  
 DEFINITION AW288561 618067F05.x1 618 - Inbred Tassel cDNA Library Zea mays cDNA, mRNA sequence.

ACCESSION AW288561  
 VERSION AW288561.1 GI:6695453  
 KEYWORDS EST.  
 SOURCE Zea mays.  
 ORGANISM Zea mays

REFERENCE 1 (bases 1 to 275)  
 AUTHORS Walbot, V.  
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Walbot V

Tomari, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y. Automated filtration-based high-throughput plasmid preparation system. *Genome Res.* 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303, 19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES	SOURCE
1. <b>High Accuracy:</b> The model achieves a high accuracy rate, consistently performing well across various datasets and tasks.	1. <b>High Accuracy:</b> The model achieves a high accuracy rate, consistently performing well across various datasets and tasks.
2. <b>Scalability:</b> The model is designed to handle large-scale data and complex tasks, making it suitable for enterprise-level applications.	2. <b>Scalability:</b> The model is designed to handle large-scale data and complex tasks, making it suitable for enterprise-level applications.
3. <b>Interpretability:</b> The model provides clear and understandable explanations for its predictions, enhancing transparency and trust.	3. <b>Interpretability:</b> The model provides clear and understandable explanations for its predictions, enhancing transparency and trust.
4. <b>Robustness:</b> The model is highly robust to noise and outliers, ensuring reliable performance in real-world scenarios.	4. <b>Robustness:</b> The model is highly robust to noise and outliers, ensuring reliable performance in real-world scenarios.
5. <b>Efficiency:</b> The model is optimized for fast processing times, allowing for real-time analysis and decision-making.	5. <b>Efficiency:</b> The model is optimized for fast processing times, allowing for real-time analysis and decision-making.
6. <b>Flexibility:</b> The model can be easily adapted to different domains and tasks, providing a versatile solution for various needs.	6. <b>Flexibility:</b> The model can be easily adapted to different domains and tasks, providing a versatile solution for various needs.
7. <b>Security:</b> The model incorporates advanced security measures to protect data and ensure compliance with industry standards.	7. <b>Security:</b> The model incorporates advanced security measures to protect data and ensure compliance with industry standards.
8. <b>Integration:</b> The model seamlessly integrates with existing systems and workflows, facilitating easy adoption and implementation.	8. <b>Integration:</b> The model seamlessly integrates with existing systems and workflows, facilitating easy adoption and implementation.
9. <b>Customization:</b> The model offers a high degree of customization, allowing users to tailor the model to their specific requirements.	9. <b>Customization:</b> The model offers a high degree of customization, allowing users to tailor the model to their specific requirements.
10. <b>Support:</b> The model is backed by a dedicated support team, ensuring users receive timely assistance and guidance.	10. <b>Support:</b> The model is backed by a dedicated support team, ensuring users receive timely assistance and guidance.

```
1. 284
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="9430038P10"
/clone_1lb="RIKEN full-1
embryonic body between d
/tissue_type="embryonic
neck"
/dev_stage="12 days embr
/lab_host="DH10a"
```

```
neck": {
  "dev_stage": "12 days embryo"
  "lab_host": "DH10B"
}
```

prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in [/note=Site 1: Sall; Site 2: BamHI; CDNA library was](#)

[illegible]

cap-trapper. cDNA went through one round of normalization and subsequently enriched for full-length by subtracting to a pool = 20.0 and subtracting to a pool = 370.0 second

CO RO = 20.0 and subduction CO RO = 370.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAGATTCGAGTTAATTAAATTAATCCCCCCCCCCC  
3']. cDNA was cleaved with XhoI and BamHI." 101 C 40 g 86 t

2 38. Score 19. DB 10. Length 284.

cy 100.0%; Pred. No. 81;	
ervative 0: Mismatches	0: Indels
	0: Gaps

TGCAATGGAAGA 464

TGCAATCGAAGA 17

294 bp mRNA linear EST 22-NOV-2000  
4-010800-111-b09 HT0734 Homo sapiens CDNA, mRNA sequence.  
1 GI:11313310

Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 50 264)

1  
I  
CO 294)  
E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
H.H., Carvalho, F.P., Matsukuma, A., Bata, G.S., Simpson, D.H.,  
A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare  
ares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

sequencing of the human transcriptome with ORF expressed tags  
 1. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 Simpson A.J.G.  
 Y of Cancer Genetics  
 Institute for Cancer Research  
 Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP.

Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asmpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?i=IL5&t2=IL5-HT0734-010800-111-b09&t3=2000-08-01&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 294.  
Location/Qualifiers

## FEATURES

source

1..294  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT0734"  
/dev\_stage="Adult"  
/note="Organ: head neck; Vector: puc18; Site 1: SmaI;  
Site 2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
79 a 78 c 55 g 82 t

## BASE COUNT

ORIGIN

Query Match 2.3%; Score 19; DB 12; Length 294;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## QY

DB

728 AGATACCTGTATAACTCC 746  
|||||  
77 AGATACCTGTATAACTCC 59

## RESULT 11

BH792739/c

LOCUS

DEFINITION

BH792739 304 bp DNA linear GSS 02-APR-2002  
SALK 064976.55.25.x Arabidopsis thaliana TDNA insertion lines  
Arabidopsis thaliana genomic clone SALK\_064976.55.25.x, DNA  
sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 304)

REFERENCE

AUTHORS

Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab  
,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.,  
Zimmerman,J. and Ecker,J.R.

TITLE

JOURNAL

COMMENT

A Sequence-Indexed Library of Insertion Mutations in the  
Arabidopsis Genome  
Unpublished (2001)  
Contact: Joseph R. Ecker  
Salk Institute Genomic Analysis Laboratory (SIGNAL)  
The Salk Institute for Biological Studies  
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
Tel: 858 453 4100 x1752  
Fax: 858 538 6379  
Email: ecker@alk.edu

This is single pass sequence recovered from the left border of  
TDNA. This sequence lies within an annotated intron of At2g01470.  
Class: TDNA tagged.

## FEATURES

source

1..304  
/organism="Arabidopsis thaliana"  
/strain="Columbia 0"  
/db\_xref="taxon:3702"  
/clone="SALK\_064976.55.25.x"  
/clone\_lib="Arabidopsis thaliana TDNA insertion lines"

/note="PCR was performed on Arabidopsis thaliana lines  
each of which contains one or more TDNA insertion  
elements. The resultant fragment for each line was  
directly sequenced to determine the genomic sequence at  
the site of insertion. Details of the protocols used can  
be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)"  
110 a 73 c 45 g 76 t

## BASE COUNT

ORIGIN

Query Match 2.3%; Score 19; DB 17; Length 304;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## QY

DB

578 GTATAGGACGACATCTTAT 596  
|||||  
259 GTATAGGACGACATCTTAT 241

## RESULT 12

AQ643691

LOCUS

DEFINITION

AQ643691 332 bp DNA linear GSS 08-JUL-1999  
RPC193-ECORI-4B19.TJ RPC193-ECORI Trypanosoma brucei genomic clone  
RPC193-ECORI-4B19, DNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Trypanosoma brucei.  
Trypanosoma brucei  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
Trypanosoma.  
1 (bases 1 to 332)

REFERENCE

AUTHORS

El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,  
Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,  
Fraser,C. and Adams,M.

TITLE

JOURNAL

COMMENT

Library for gene discovery and sequence-ready map construction  
Unpublished (1999)  
Other\_GSSs: RPC193-ECORI-4B19.TV  
Contact: Najib M. El-Sayed  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: nelsayed@tigr.org

Clones and high density filters may be purchased from BACPAC  
Resources (<http://bacpac.med.buffalo.edu>). BAC end sequences search  
page: <http://www.tigr.org/tdb/mdb/tbdb/>.  
Seq primer: SP6  
Class: BAC ends.

## FEATURES

source

Location/Qualifiers  
1..332  
/organism="Trypanosoma brucei"  
/strain="TREU927/4 GUTat 10.1"  
/db\_xref="taxon:5691"  
/clone="RPC193-ECORI-4B19"  
/clone\_lib="RPC193-ECORI"  
/note="Vector: pBACe3.6; Site 1: Eco RI; Site 2: Eco RI;  
Constructed for The Institute for Genomic Research by  
Bohui Zhao in Peter de Jong's laboratory (Roswell Park  
Cancer Institute, Buffalo, NY). Briefly, Trypanosoma  
brucei TREU927/4 GUTat 10.1 agarose embedded DNA was  
partially digested with a combination of Eco RI and Eco RI  
methylase (RPC193-ECORI segment) or Dpn II (RPC193-DpnII  
segment). High molecular weight fragments were ligated in  
pBACe3.6 vector digested with Eco RI or Bam HI,  
respectively. The average insert size is 141 Kb. Total  
coverage (both segments): > 90 X the haploid  
non-minichromosomal genome."  
53 a 85 c 99 g 95 t

## BASE COUNT

ORIGIN

Query Match 2.3%; Score 19; DB 17; Length 332;

Best Local Similarity 100.0%; Pred. No. 80;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 822 TGGATGAGGTTCACTTC 840  
|||||  
Db 240 TGGATGAGGTTCACTTC 258  
|||||

RESULT 13  
Z39359  
LOCUS HSC18B032 normalized infant brain cDNA Homo sapiens CDNA clone  
DEFINITION c-18b03 3', mRNA sequence.

ACCESSION Z39359  
VERSION Z39359.1 GI:562551  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 340)  
AUTHORS Autray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes  
M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,  
Mitchell,H., Mariage-Samson,R., Pietu,G., Pouillot,Y.,  
Sebastiani-Kabackhis,C. and Tessier,A.  
TITLE IMAGE: molecular integration of the analysis of the human genome  
and its expression  
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)  
MEDLINE 95277534  
COMMENT Contact: Genethon  
Genexpross-Genethon  
Genethon Centre de recherche sur le Genome Humain  
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE  
Tel: 33169472800  
Fax: 33160778698  
Email: genexpross@genethon.fr

Single read, 24 t removed at sequence 5'end  
Genexpress library\_idt: C; Genexpress\_sequence\_idt: alc-18b03  
Seq primer: (-21)M13 universal.  
FEATURES  
source  
1..340  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clones="c-18b03"  
/clone\_lib="normalized infant brain cDNA"  
/sex="Female"  
/tissue\_type="total brain"  
/dev\_stage="3 months old"  
/note="Organ: brain; Vector: lafmid BA; Site 1: HindIII;  
Site 2: NotI; sex:Female; dev stage=3 months old;  
isolates:muscular atrophy patient; tissue type=total brain  
; total mRNA was oligo-(dt) primed and directionally  
cloned 5' -> 3' into the HindIII -> NotI sites of the  
lafmid BA vector. Clone library from B.Souares, Psychiatry  
Dept. Columbia University, USA. Normalization\_method:  
Bento Soares, P.N.A.S in press"

BASE COUNT 116 a 55 c 49 g 119 t 1 others  
ORIGIN

Query Match 2.3%; Score 19; DB 14; Length 340;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 591 TCTTATCACTATTTTAAA 609  
|||||  
Db 264 TCTTATCACTATTTTAAA 282  
|||||

RESULT 14  
BF590417  
LOCUS BF590417 363 bp mRNA linear EST 12-DEC-2000  
DEFINITION na37901.x1 NCI\_CGAP\_Kid11 Homo sapiens CDNA clone IMAGE:3258792  
3', mRNA sequence.

ACCESSION BF590417  
VERSION BF590417.1 GI:11682741  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 363)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
info@image.llnl.gov

Seq primer: -40UP from Gibco  
High quality sequence stop: 327.  
Location/Qualifiers  
1..363  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clones="IMAGE:3258792"  
/clone\_lib="NCI\_CGAP\_Kid11"  
/lab\_hosts="DH10B"  
/note="Organ: kidney; Vector: pTT73D-Pac (Pharmacia) with  
a modified polylinker; Site 1: Not I; Site 2: Eco RI;  
Plasmid DNA from the normalized library NCI\_CGAP Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneIDs 1322376-1323911, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo."

BASE COUNT 127 a 57 c 53 g 126 t  
ORIGIN

Query Match 2.3%; Score 19; DB 12; Length 363;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 591 TCTTATCACTATTTTAAA 609  
|||||  
Db 264 TCTTATCACTATTTTAAA 282  
|||||

RESULT 15  
AV414017  
LOCUS AV414017 434 bp mRNA linear EST 23-MAY-2000  
DEFINITION japonicus cDNA clone MMW23901\_r 5', mRNA sequence.

ACCESSION AV414017  
VERSION AV414017.1 GI:7743193  
KEYWORDS EST.  
SOURCE Lotus japonicus  
ORGANISM Lotus japonicus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;  
Lotus.  
REFERENCE 1 (bases 1 to 434)  
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.  
TITLE Generation of 7137 non-redundant expressed sequence tags from a  
legume, Lotus japonicus  
JOURNAL DNA Res. 7 (2), 127-130 (2000)  
MEDLINE 20277479

## COMMENT

Contact: Yasukazu Nakamura  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 252-0812, Japan  
Email: ynakamu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

## FEATURES

source

1..434  
/organism="Lotus japonicus"  
/db\_xref="taxon:34305"  
/clone="MM239G01.r"  
/clone\_lib="Lotus japonicus young plants (two-week old)"  
/dev\_stage="young plants (two-week old)"  
/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
XhoI; isolate=miyakojima MG-20"  
144 a 87 c 126 g 77 t

BASE COUNT  
ORIGIN

Query Match 2.3%; Score 19; DB 10; Length 434;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 TTGAAGCTGCATATCAACA 374  
|||||  
DB 106 TTGAAGCTGCATATCAACA 124

## RESULT 16

AA730342

LOCUS

DEFINITION  
nw46b11.s1 NCI CGAP Ew1 Homo sapiens cDNA clone IMAGE:1249629  
similar to contains Alu repetitive element; contains element L1  
repetitive element ;, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS

TITLE

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

## JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,  
Ph.D.

CDNA Library Preparation: David B. Krizman, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 658 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham.

## FEATURES

source

1..441  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1249629"  
/clone\_lib="NCI CGAP\_Ew1"  
/tissue\_type="Ewing's sarcoma"  
/lab\_host="DH10B"  
/note="Vector: pAMP10; mRNA made from Ewing's sarcoma,  
made by oligo-dT priming. Non-directionally cloned.  
Size selected on agarose gel, average insert size 600 bp.  
Reference: Krizman et al. (1996) Cancer Research  
56:5380-5383."

BASE COUNT  
ORIGIN

94 a 85 c 76 g 186 t

## Query Match

2.3%; Score 19; DB 9; Length 441;

Best Local Similarity 100.0%; Pred. No. 79;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 TTTTCAGATCTCTGTAGGT 87  
|||||  
DB 152 TTTTCAGATCTCTGTAGGT 170

## RESULT 17

AA946673

LOCUS

DEFINITION  
AA946673.1 NCI CGAP GCBI Homo sapiens cDNA clone IMAGE:1372154 3',  
mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS

TITLE

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

## JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,  
Ph.D., Gerald Marti, M.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1105 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 452.

Location/Qualifiers

1..452

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1372154"

/clone\_lib="NCI CGAP GCBI"

/tissue\_type="germinal center B cell"

/lab\_host="DH10B"

/note="Vector: pTT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD-),  
provided by Dr. Louis M. Staudt (NCI). Dr. David Allman  
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
primed with a Not I - oligo(dT) primer

15'-GGTACCAATCTGAAGCGCGCCCTCATTTTTTTTTTTT-3',  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pTT73 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

## FEATURES

source

154 a 68 c 62 g 168 t  
BASE COUNT  
ORIGIN



```

Query Match      2.3%; Score 19; DB 9; Length 452;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 591 TCTATCACTATTTTAA 609
Db 279 TCTATCACTATTTTAA 297

RESULT 18
R60488
LOCUS R60488 454 bp mRNA linear EST 24-MAY-1995
DEFINITION YH13a11.s1 Soares infant brain INIB Homo sapiens cDNA clone
IMAGE:42836 3', mRNA sequence.
ACCESSION R60488
VERSION R60488.1 GI:831183
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 454)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, J., Soares, M., Tan, F., Trevaskis, E., Waterston
, R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Contact: Wilson RK
Unpublished (1995)
TITLE Washington University School of Medicine
JOURNAL 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
COMMENT Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 305
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: SP6
High quality sequence stop: 305.
FEATURES
Location/Qualifiers
source
1..454
/organism="Homo sapiens"
/db_xref="GDB:415377"
/db_xref="taxon:9606"
/clone="IMAGE:42836"
/clone_lib="Soares infant brain INIB"
/sex="female"
/dev_stages="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: whole brain; Vector: Lfamid BA; Site 1: Not
I; Site 2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5',
AATCGGAGAGATTCGGCCGACAGATTTTCTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lfamid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 144 a 69 c 75 g 161 t
ORIGIN
Query Match      2.3%; Score 19; DB 14; Length 454;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 591 TCTATCACTATTTTAA 609
Db 280 TCTATCACTATTTTAA 298

RESULT 19
AU090412/c

```

```

LOCUS AU090412 459 bp mRNA linear EST 30-JUL-2002
DEFINITION AU090412 Paralicthys olivaceus kidney injected with peptidoglycan
Paralicthys olivaceus cDNA clone M200, mRNA sequence.
ACCESSION AU090412
VERSION AU090412.1 GI:7608628
KEYWORDS EST.
SOURCE Baetard hallbut.
ORGANISM Paralicthys olivaceus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
TITLE Acanthomorpha; Acanthopterygii; Percormorpha; Pleuronectiformes;
Pleuronectoidae; Paralicthyidae; Paralicthys.
1 (bases 1 to 459)
Sakai, M., Kono, T. and Oshikawa, T.
The EST analysis of kidney cells Japanese flounder, Paralicthys
olivaceus, injected with peptidoglycan
Unpublished (2000)
JOURNAL Contact: Masahiro Sakai
COMMENT Faculty of Agriculture
Miyazaki University
1-1 nishi gakuenkibanadai, Miyazaki, Miyazaki 889-2192, Japan
Email: m.sakai@cc.miyazaki-u.ac.jp.
FEATURES
Location/Qualifiers
source
1..459
/organism="Paralicthys olivaceus"
/db_xref="taxon:8255"
/clone="M200"
/clone_lib="Paralicthys olivaceus Kidney injected with
peptidoglycan"
/tissue_type="kidney"
/notes="common name: Japanese flounder ; injected with
peptidoglycan"
BASE COUNT 110 a 115 c 147 g 87 t
ORIGIN
Query Match      2.3%; Score 19; DB 9; Length 459;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 CTCGCTGAGAGAACTCCT 183
Db 248 CTCGCTGAGAGAACTCCT 230

RESULT 20
AQ455854/c
LOCUS AQ455854 462 bp DNA linear GSS 21-APR-1999
DEFINITION HS_5151_A1 G02_SP6E RPCI-11 Human Male BAC Library Homo sapiens
Genomic clone Plate=727 Col=3 Row=M, DNA sequence.
ACCESSION AQ455854
VERSION AQ455854.1 GI:4591189
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 462)
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D., and
Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
JOURNAL 9380589
MEDLINE
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong

```

(pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources ([http://bacpac.med.buffalo.edu/ordering\\_bac.htm](http://bacpac.med.buffalo.edu/ordering_bac.htm)) or from Research Genetics (<http://www.htsc.washington.edu>) Plate: 727 row: M column: 3 Seq primer: SP6 Class: BAC ends High quality sequence stop: 462.

## FEATURES

source

Location/Qualifiers  
1..462  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate-727 Col-3 Row-M"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"  
79 a 104 c 100 g 169 t 10 others

BASE COUNT  
ORIGIN

Query Match 2.3%; Score 19; DB 17; Length 462;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 724 GAGAAGATACCTGTGAATAA 742  
|||||  
Db 62 GAGAAGATACCTGTGAATAA 44

RESULT 21

AZ697618/c

LOCUS

DEFINITION RPCI-23-220C15, TU RPCI-23 Mus musculus genomic clone RPCI-23-220C15  
DNA sequence.

ACCESSION

AZ697618

VERSION

AZ697618.1

KEYWORDS

GSS.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)  
Other\_GSSs: RPCI-23-220C15.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 220 row: C column: 15  
Seq primer: SP6  
Class: BAC ends.

## FEATURES

source

Location/Qualifiers  
1..464  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-220C15"  
/clone\_lib="RPCI-23"

/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."  
169 a 86 c 97 g 112 t

Query Match 2.3%; Score 19; DB 17; Length 464;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 CATTGATTTTCAAAATAAC 297  
|||||  
Db 20 CATTGATTTTCAAAATAAC 2

RESULT 22

BI322121

LOCUS

DEFINITION BI322121 467 bp mRNA linear EST 29-NOV-2001  
ID: Gm-cl051-8260 5' similar to SW:IF3A\_TOBAC Q40554 EUKARYOTIC  
TRANSLATION INITIATION FACTOR 3 SUBUNIT 10 ; mRNA sequence.

ACCESSION

BI322121

VERSION

BI322121.1

KEYWORDS

EST.

SOURCE

ORGANISM

Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

Glycine.

REFERENCE

1 (bases 1 to 467)

AUTHORS

Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna

,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,

Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers

,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk

,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann

,R., Waterston,R. and Wilson,R.

Public Soybean EST Project

Unpublished (1999)

CONTACT: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)

This clone is available through: RegGen, Invitrogen Corp. 2130

South Memorial Parkway Huntsville, AL 35801 For further information

call: (800)-533-4363 or contact via email: [ccu@resgen.com](mailto:ccu@resgen.com)

Seq primer: -40RP from Gibco

High quality sequence stop: 383.

Location/Qualifiers

1..467

/organism="Glycine max"

/db\_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl051-8260"

/clone\_lib="Gm-cl051"

/issue\_type="floral meristematic mRNA"

/lab\_host="DH10B"

/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:

XhoI; The cDNA library was constructed from floral

meristematic mRNA provided by Dr. Halina Knap of Clemson

University. Complementary DNA was synthesized from mRNA

using a primer consisting of a poly(dT) sequence with a

XhoI restriction site. EcoRI adapters were ligated to the

blunt-ended cDNA fragments followed by XhoI digestion. The

cDNA fragments were directionally cloned into the

EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."

BASE COUNT 179 a 64 c 149 g 75 t

Query Match 2.3%; Score 19; DB 13; Length 467;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 TTGAAGCTGCATATCAACA 374

Db 25 TTGAAGCTGCATATCAACA 43

RESULT 23

LOCUS BG041760

DEFINITION BG041760 472 bp mRNA linear EST 06-DEC-2001  
SV38f08.y1 Gm-cl057 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
Gm-cl057-1816 5' similar to SW:IF3A.TOBAC Q40554 EUKARYOTIC  
TRANSLATION INITIATION FACTOR 3 SUBUNIT 10 ; mRNA sequence.

ACCESSION BG041760

VERSION BG041760.1 GI:12488115

KEYWORDS EST.

SOURCE soybean.

ORGANISM Glycine max

REFERENCE 1 (bases 1 to 472)  
Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna  
A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,  
Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers  
Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk  
R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann  
R., Waterston,R. and Wilson,R.  
Public Soybean EST Project  
Unpublished (1999)

TITLE

JOURNAL

COMMENT

Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

Trace considered overall poor quality This clone is available

through: Resgen, Invitrogen Corp. 2130 South Memorial Parkway

Huntsville, AL 35801 For further information call: (800)-533-4363

or contact via email: ccu@resgen.com

High quality sequence stop: 1.

Location/Qualifiers

1. .472

/organism="Glycine max"

/db\_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl057-1816"

/clone\_lib="Gm-cl057"

/tissue\_type="Degenerating cotyledons, 2 week old seedling"

/lab\_host="DH10B"

/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from mRNA isolated from degenerating cotyledons of 2 week old seedlings from PI489916. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."

FEATURES

source

Query Match 2.3%; Score 19; DB 14; Length 472;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 591 TCTTATCACTATTTTAA 609

Db 279 TCTTATCACTATTTTAA 297

BASE COUNT 175 a 77 c 130 g 90 t

ORIGIN

Query Match 2.3%; Score 19; DB 12; Length 472;

Best Local Similarity 100.0%; Pred. No. 79;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 TTGAAGCTGCATATCAACA 374

Db 165 TTGAAGCTGCATATCAACA 183

RESULT 24

LOCUS R37263

DEFINITION R37263 472 bp mRNA linear EST 04-MAY-1995  
YF55e05.s1 Soares infant brain IN1B Homo sapiens cDNA clone  
IMAGE:25776 3', mRNA sequence.

ACCESSION R37263

VERSION R37263.1 GI:794719

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 472)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman  
M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,  
Rifkin,L., Rohlfing,T., Soares,M., Tan,P., Trevasakis,E., Waterston  
R., Williamson,A., Wohlmann,P. and Wilson,R.  
The WashU-Merck EST Project  
Unpublished (1995)

JOURNAL

COMMENT

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 1272

High quality sequence stops: 313 Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1272 Std Error: 0.00

Seq primer: -21ml3

High quality sequence stop: 313.

FEATURES

source

1. .472

/organism="Homo sapiens"

/db\_xref="GDB:398123"

/db\_xref="taxon:9606"

/clones="IMAGE:25776"

/clone\_lib="Soares infant brain IN1B"

/sex="female"

/dev\_stage="73 days post natal"

/lab\_host="DH10B (ampicillin resistant)"

/note="Organ: whole brain; Vector: Lfamid BA; Site 1: Not

I; Site 2: Hind III; 1st strand cDNA was primed with a Not

I - oligo(dT) primer (5'

AACCTGGAAGATTGCGCGCAGGAATTTTTTTTTTTT 3');

double-stranded cDNA was ligated to Hind III adaptors

(Pharmacia), digested with Not I and directionally cloned

into the Not I and Hind III sites of the Lfamid BA vector.

Library went through one round of normalization. Library

constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 142 a 75 c 81 g 166 t

ORIGIN

```

RESULT 25
R43173
LOCUS
DEFINITION
Yg18c04.s1 Soares infant brain INIB Homo sapiens cDNA clone
IMAGE:32749 3', mRNA sequence.
ACCESSION
R43173
VERSION
R43173.1 GI:825399
KEYWORDS
EST.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 473)
AUTHORS
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston
R., Williamson, A., Wohldmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
JOURNAL
On May 8, 1995 this sequence version replaced gi:801397.
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1314
High quality sequence stops: 305 Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1314 Std Error: 0.00
Seq primer: Promega -21ml3
High quality sequence stop: 305.
FEATURES
Location/Qualifiers
1..473
/organism="Homo sapiens"
/db_xref="GB:405096"
/db_xref="taxon:9606"
/clone="IMAGE:32749"
/clone_lib="Soares infant brain INIB"
/sex="Female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/Note="Organ: whole brain; Vector: Latmid BA; Site:1: Not
I; Site 2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5',
AACTGGAAGAATTCGGCGCCGAGGAATTTTTTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Latmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Patima Bonaldo."
BASE COUNT
134 a 77 c 87 g 168 t 7 others
ORIGIN
Query Match 2.3%; Score 19; DB 14; Length 473;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 591 TCTATCACTATTTTAAA 609
|||||
DB 281 TCTATCACTATTTTAAA 299

RESULT 26
AQ644492
LOCUS
DEFINITION
RPC193-EcoRI-3F9, TP RPC193-EcoRI Trypanosoma brucei genomic clone
ACCESSION
AQ644492
VERSION
AQ644492.1 GI:5121202
KEYWORDS
EST.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 493)
AUTHORS
El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,
Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.
Use of BAC end sequences from Trypanosoma brucei GUTat 10.1 RPCI-93
Library for gene discovery and sequence-ready map construction
Unpublished (1999)
JOURNAL
Other GSSs: RPC193-EcoRI-3F9-TV
CONTACT: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones and high density filters may be purchased from BACPAC
Resources (http://bacpac.med.buffalo.edu). BAC end sequences search
page: http://www.tigr.org/tdb/mbd/tbdb/.
Seq primer: SP6
Class: BAC ends.
FEATURES
Location/Qualifiers
1..493
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="RPC193-EcoRI-3F9"
/clone_lib="RPC193-EcoRI"
/Note="Vector: pBACE3.6; Site:1: Eco RI; Site 2: Eco RI;
Constructed for The Institute for Genomic Research by
Bohui Zhao in Pieter de Jong's laboratory (Roswell Park
Cancer Institute, Buffalo, NY). Briefly, Trypanosoma
brucei TREU927/4 GUTat 10.1 agarose embedded DNA was
partially digested with a combination of Eco RI and Eco RI
methyase (RPC193-EcoRI segment) or Dpn II (RPC193-DpnII
segment). High molecular weight fragments were ligated in
pBACE3.6 vector digested with Eco RI or Bam HI,
respectively. The average insert size is 141 Kb. Total
coverage (both segments): > 90 X the haploid
non-minichromosomal genome."
BASE COUNT
100 a 129 c 136 g 128 t
ORIGIN
Query Match 2.3%; Score 19; DB 17; Length 493;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 822 TGAATGAGGTTACCTTC 840
|||||
DB 331 TGAATGAGGTTACCTTC 349

RESULT 27
AQ071633
LOCUS
DEFINITION
WT94d10.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2515123 3',
mRNA sequence.
ACCESSION
AQ071633
VERSION
AQ071633.1 GI:6026631
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 494)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL

```

```

KEYWORDS
GSS.
SOURCE
Trypanosoma brucei.
ORGANISM
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE
1 (bases 1 to 493)
AUTHORS
El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,
Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.
Use of BAC end sequences from Trypanosoma brucei GUTat 10.1 RPCI-93
Library for gene discovery and sequence-ready map construction
Unpublished (1999)
JOURNAL
Other GSSs: RPC193-EcoRI-3F9-TV
CONTACT: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones and high density filters may be purchased from BACPAC
Resources (http://bacpac.med.buffalo.edu). BAC end sequences search
page: http://www.tigr.org/tdb/mbd/tbdb/.
Seq primer: SP6
Class: BAC ends.
FEATURES
Location/Qualifiers
1..493
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="RPC193-EcoRI-3F9"
/clone_lib="RPC193-EcoRI"
/Note="Vector: pBACE3.6; Site:1: Eco RI; Site 2: Eco RI;
Constructed for The Institute for Genomic Research by
Bohui Zhao in Pieter de Jong's laboratory (Roswell Park
Cancer Institute, Buffalo, NY). Briefly, Trypanosoma
brucei TREU927/4 GUTat 10.1 agarose embedded DNA was
partially digested with a combination of Eco RI and Eco RI
methyase (RPC193-EcoRI segment) or Dpn II (RPC193-DpnII
segment). High molecular weight fragments were ligated in
pBACE3.6 vector digested with Eco RI or Bam HI,
respectively. The average insert size is 141 Kb. Total
coverage (both segments): > 90 X the haploid
non-minichromosomal genome."
BASE COUNT
100 a 129 c 136 g 128 t
ORIGIN
Query Match 2.3%; Score 19; DB 17; Length 493;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 822 TGAATGAGGTTACCTTC 840
|||||
DB 331 TGAATGAGGTTACCTTC 349

RESULT 27
AQ071633
LOCUS
DEFINITION
WT94d10.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2515123 3',
mRNA sequence.
ACCESSION
AQ071633
VERSION
AQ071633.1 GI:6026631
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 494)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL

```

## COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
www.bto.llnl.gov/bbrp/image/image.html  
Seq primer: -40UP from Gibco  
High quality sequence stop: 455.

## FEATURES

source

Location/Qualifiers  
1. .494  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2515123"  
/clone\_lib="NCI-CGAP\_GCs"  
/tissue\_type="pooled germ cell tumors"  
/lab\_host="DH10B"

/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP\_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

172 a 79 c 76 g 167 t

ORIGIN

Query Match 2.3%; Score 19; DB 10; Length 494;

Best Local Similarity 100.0%; Pred. No. 79;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

591 TCTTATCACTATTTTAA 609

Db

245 TCTTATCACTATTTTAA 263

RESULT 28

AW508943

LOCUS

AW508943 513 bp mRNA linear EST 03-DEC-2001  
s138c03.v1 Gm-r1030 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
Gm-r1030-1181 5' similar to FR:Q40554 Q40554 PNLA-35. ;, mRNA  
sequence.

ACCESSION

AW508943.1 GI:7147021

VERSION

AW508943.1

KEYWORDS

EST.

SOURCE

soybean.

ORGANISM

Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Glycine.

REFERENCE

1 (bases 1 to 513)

AUTHORS

Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna  
A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,  
Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers  
Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk  
R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann  
R., Waterston, R. and Wilson, R.

TITLE

Public Soybean EST Project

JOURNAL

Unpublished (1999)

COMMENT

Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

## FEATURES

source

Location/Qualifiers  
1. .513  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-r1030-1181"  
/clone\_lib="Gm-r1030"  
/lab\_host="DH10B"

/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This cDNA library was constructed from mRNA isolated from immature cotyledons of greenhouse grown plants (individual seed fresh weight of 100-300mg). The library was prepared using the Life Technologies pSuperScript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly(dT) sequence with a NotI restriction site. SalI linkers adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E. coli ElectroMax DH10B host cells. This library was constructed by Dr. Lila Vodkin and Dr. Anu Khanna. Note that Gm-r1030 is a re-track of Gm-cl007."

BASE COUNT

196 a 76 c 152 g 87 t 2 others

ORIGIN

Query Match 2.3%; Score 19; DB 10; Length 513;

Best Local Similarity 100.0%; Pred. No. 79;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

356 TTGAAGCTGCATATCAACA 374

Db

149 TTGAAGCTGCATATCAACA 167

RESULT 29

AW648727

LOCUS

AW648727 530 bp mRNA linear EST 18-MAY-2001  
EST327097 tomato germinating seedlings, TAMU lycopersicon  
esculentum cDNA clone CLEISG14 5', mRNA sequence.

ACCESSION

AW648727.1 GI:7409881

VERSION

AW648727

KEYWORDS

EST.

SOURCE

tomato.

ORGANISM

Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.

REFERENCE

1 (bases 1 to 530)

AUTHORS

Alcala, J., Vrebalov, J., White, R., van der Hoeven, R.S., Holt, I.E.,  
Liang, F., Hansen, T.S., Craven, M.B., Bowman, C.L., Ronning, C.M.,  
Niernman, W., Fraser, C.M., Giovannoni, J.J., Martin, G.B. and Tanksley  
S.D.

TITLE

Generation of ESTs from germinating tomato seed

JOURNAL

Unpublished (2000)

COMMENT

Contact: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

FEATURES

source

Location/Qualifiers

1. .530

/organism="Lycopersicon esculentum"

/cultivar="TA496"

/db\_xref="taxon:4081"

/clone="CLEISG14"

/clone\_lib="tomato germinating seedlings, TAMU"

```

/tissue_type="whole seedlings"
/dev_stages=7 days post imbibition"
/notes="Vector: pBlueScript SK(-); Site 1: EcoR1; Site 2:
Xho1; 7 days post imbibition on water-agar. Mixed stage
whole germinating seedlings from seed coat emergence up
to two centimeters in seeds not showing obvious signs of
germination were discarded."
BASE COUNT 175 a 108 c 77 g 169 t 1 others
ORIGIN

Query Match 2.3%; Score 19; DB 10; Length 530;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 CAACAATTTAATCCAAAAA 388
|||||
Db 232 CAACAATTTAATCCAAAAA 250

RESULT 30
AW650242 535 bp mRNA linear EST 18-MAY-2001
LOCUS EST328696 tomato germinating seedlings, TAMU Lycopersicon
DEFINITION esculentum cDNA clone cLE112M16 5', mRNA sequence.
ACCESSION AW650242
VERSION AW650242.1 GI:7411480
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 535)
AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S, Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Giovannoni,J.J., Martin,G.B. and Tanksley
,S.D.
TITLE Generation of ESTs from germinating tomato seed
JOURNAL Unpublished (2000)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
Location/Qualifiers
source
1..535
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLE112M16"
/clone_lib="tomato germinating seedlings, TAMU"
/tissue_type="whole seedlings"
/dev_stages="7 days post imbibition"
/notes="Vector: pBlueScript SK(-); Site 1: EcoR1; Site 2:
Xho1; 7 days post imbibition on water-agar. Mixed stage
whole germinating seedlings from seed coat emergence up
to two centimeters in seeds not showing obvious signs of
germination were discarded."
BASE COUNT 169 a 115 c 80 g 170 t 1 others
ORIGIN

Query Match 2.3%; Score 19; DB 10; Length 535;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 CAACAATTTAATCCAAAAA 388
|||||
Db 387 CAACAATTTAATCCAAAAA 405

RESULT 31

```

```

AW206419 543 bp mRNA linear EST 02-DEC-1999
LOCUS UI-H-BII-afg-d-02-0-UI.s1 NCI_CGAP_Sub3 Homo sapiens CDNA clone
DEFINITION IMAGE:2721818 3', mRNA sequence.
ACCESSION AW206419
VERSION AW206419.1 GI:6505915
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 543)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LINL at:
www-bio.lnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=Yes.
FEATURES
Location/Qualifiers
source
1..543
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2721818"
/clone_lib="NCI_CGAP_Sub3"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The
NCI_CGAP Sub3 library is a subtracted library derived from
the NCI_CGAP Sub1 library, which is a subtracted library
derived from Bi. Bi constitutes a mixture of 21
normalized or subtracted NCI_CGAP libraries: NCI_CGAP_Co4
, NCI_CGAP_Pr22, NCI_CGAP_Pr28, NCI_CGAP_Co10,
NCI_CGAP_Kid3, NCI_CGAP_Kid5, NCI_CGAP_Kid12,
NCI_CGAP_Kid3, NCI_CGAP_Kid11, NCI_CGAP_Lym2,
NCI_CGAP_Br2, NCI_CGAP_Co8, NCI_CGAP_CLL1, NCI_CGAP_Lei2,
NCI_CGAP_Brn23, NCI_CGAP_Lu5, NCI_CGAP_Lu24,
NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6,
NCI_CGAP_Brn25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with
a driver whose composition is detailed below:
NCI_CGAP_Kid3 pool 1 LLAM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clones 132376-132391), 1456008-1456775
, 1500552-1502855); NCI_CGAP_Kid5 pool 1 LLAM 3338-3342
, 3722-3725, 3776-3778 (IMAGE Clones 1323912-1325831,
1471368-1472903, 1492104-1493255); NCI_CGAP_Lu5 pool 1
LLAM 3575-3582, 3851-3854 (IMAGE Clones 1414920-1417991,
1529094-1522459); NCI_CGAP_GC4 pool 1 LLAM 3164-3167,
3716-3720, 3733-3735 (IMAGE Clones 1257096-1258631,
1469064-1470983, 1475592-1476743); NCI_CGAP_Pr22 pool 1
LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clones
985608-986759, 1101192-1101959, 1217928-1220615);
NCI_CGAP_Co10 pool 1 LLAM 2644-2653, 2871-2872 (IMAGE
Clones 1057416-1061255, 114584-1145351). Subtraction
was performed as previously described (Bonaldi, Lennon &
Soares (1996): Normalization and Subtraction: Two
Approaches To Facilitate Gene Discovery. Genome Research
6, 791-806.
TAG LIB=NCI_CGAP_GC4
TAG_TISSUE=germ cell
TAG_SEQ=AAATC"
BASE COUNT 180 a 79 g 199 t
ORIGIN

```

```

Query Match      2.3%; Score 19; DB 10; Length 543;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 591 TCTTATCACTATTTTAA 609
Db 262 TCTTATCACTATTTTAA 280

RESULT 32
LOCUS H05653
DEFINITION Y175b11.s1 Soares infant brain 1N1B Homo sapiens cDNA clone
IMAGE:43770 3', mRNA sequence.
ACCESSION H05653
VERSION H05653.1 GI:869205
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 550)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
R., Williamson,A., Woldmann,P. and Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1378
High quality sequence stops: 283
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1378 Std Error: 0.00
Seq primer: Promega -21ml3
High quality sequence stop: 283.
Location/Qualifiers
1..550
/organism="Homo sapiens"
/db_xref="GDB:416311"
/db_xref="taxon:9606"
/clone="IMAGE:43770"
/clone_lib="Soares infant brain 1N1B"
/sex="female"
/dev_stages="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="organ: whole brain; Vector: Lfamid BA; Site 1: Not
I; Site 2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(GT) primer [5'
AACGTGAGAAATTCGCGCCGACGAAATTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lfamid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Patima Bonaldo."
BASE COUNT 166 a 85 c 93 g 191 t 15 others
ORIGIN

Query Match      2.3%; Score 19; DB 14; Length 550;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 591 TCTTATCACTATTTTAA 609
Db 279 TCTTATCACTATTTTAA 297

```

```

RESULT 33
LOCUS BP017649/c
DEFINITION BP017649 Nori Satoh unpublished cDNA library, young adult Ciona
intestinalis cDNA clone ciad69g12 5', mRNA sequence.
ACCESSION BP017649
VERSION BP017649.1 GI:19509126
KEYWORDS EST.
SOURCE Ciona intestinalis.
ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Cionidae; Ciona.
1 (bases 1 to 554)
AUTHORS Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
TITLE Expressed genes in Ciona intestinalis
JOURNAL Unpublished (2000)
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
1..554
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone="ciad69g12"
/clone_lib="Nori Satoh unpublished cDNA library, young
adult"
/tissue_type="whole animal"
/dev_stage="young adult"
/note="Vector: pBluescript SK"
BASE COUNT 182 a 101 c 110 g 157 t 4 others
ORIGIN

Query Match      2.3%; Score 19; DB 14; Length 554;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 105 CAAAGAGGCTTCTACATT 123
Db 549 CAAAGAGGCTTCTACATT 531

RESULT 34
LOCUS TA72C09P/c
DEFINITION T. brucei sheared genomic DNA clone 72c09, forward sequence,
genomic survey sequence.
ACCESSION AL457099
VERSION AL457099.1 GI:11859458
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 558)
AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
Melville,S.E., Rajandream,M.A. and Barrell,B.G.
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In

```

Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999).  
 Email: nelsayed@igr.org  
 Details of T. brucei sequencing at the Sanger Centre are available at [http://www.sanger.ac.uk/Projects/T\\_brucei/](http://www.sanger.ac.uk/Projects/T_brucei/).

## FEATURES

source

1..558  
 /organism="Trypanosoma brucei"  
 /strain="TREU927"  
 /db\_xref="taxon:5691"  
 /clone="72C09"

BASE COUNT 145 a 167 c 133 g 113 t

## ORIGIN

Query Match 2.3%; Score 19; DB 17; Length 559;  
 Best Local Similarity 100.0%; Pred. No. 78;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 822 TCGAATGAGGTTCACTTC 840

Db 414 TCGAATGAGGTTCACTTC 396

## RESULT 35

BE921543

## LOCUS

DEFINITION BE921543 559 bp mRNA linear EST 02-OCT-2000  
 CSTB14110 5' sequence, mRNA sequence.

## ACCESSION

BE921543

## VERSION

BE921543.1

## KEYWORDS

EST.

## SOURCE

POTATO.

## ORGANISM

Solanum tuberosum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 van der Hoeven, R.S., Berzerides, J., Holt, I.E., Liang, F., Cho, J.,  
 Utterbeck, T., Hansen, C.L., Doan, B., Bougri, O., Buell, C.R., Renning  
 C.M., Fry, W.E., Tanksley, S.D. and Baker, B.  
 Generation of ESTs from potato leaves and petioles  
 Unpublished (2000)

## COMMENT

Contact: Cathy Renning  
 The Institute for Genomic Research  
 For clone request: please contact Research Genetics, Libraries  
 Division tel 1-800-711-6195, email [cdna@resgen.com](mailto:cdna@resgen.com).

## FEATURES

source

1..559  
 /organism="Solanum tuberosum"  
 /cultivar="Kennebec"  
 /db\_xref="taxon:4113"  
 /clone="cSTB14110"  
 /clone\_lib="potato leaves and petioles"  
 /tissue\_type="leaflets and petioles"  
 /dev\_stage="8 weeks old plants"  
 /lab\_host="SOLR"  
 /note="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:  
 XhoI; Tissue was supplied by Dr. Fry (Cornell University).  
 Leaflets and petioles were isolated from 8 week old  
 greenhouse grown plants. The plants were watered and  
 fertilized freely. The tissue was immediately frozen in  
 liquid nitrogen."

BASE COUNT 156 a 107 c 134 g 162 t

## ORIGIN

Query Match 2.3%; Score 19; DB 12; Length 559;  
 Best Local Similarity 100.0%; Pred. No. 78;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 198 TTCTCTCACTAAAAAGTT 216

Db 10 TTCTCTCACTAAAAAGTT 28

## RESULT 36

AI667027

## LOCUS

DEFINITION AI667027 561 bp mRNA linear EST 07-JUN-2001  
 IMAGE:3722356 5', mRNA sequence.

## ACCESSION

AI667027

## VERSION

AI667027.1

## KEYWORDS

EST.

## SOURCE

zebrafish.

## ORGANISM

Danio rerio

## REFERENCE

AUTHORS

1 (bases 1 to 561)

; Cyprinidae; Danio.

Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy

, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wyllie, T., Underwood

, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,

Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,

Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.

and Wilson, R.

WashU Zebrafish EST Project 1998

Unpublished (1998)

Contact: Stephen L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: [zbrfish@wustl.edu](mailto:zbrfish@wustl.edu)

CDNA Library Preparation: Matthew Clark. CDNA Library Arrayed by:

Matthew Clark. DNA Sequencing by: Washington University Genome

Sequencing Center Clone distribution: Genome Systems, St. Louis,

Missouri (web address: [www.genomesystems.com](http://www.genomesystems.com)) (email contact:[info@genomesystems.com](mailto:info@genomesystems.com)) and Research Genetics, Huntsville, Alabama(web address: [www.resgen.com](http://www.resgen.com)) (email contact: [info@resgen.com](mailto:info@resgen.com)) and

ResourcenZentrumPrimarDatenbank, Berlin, Germany (web address:

[www.rzpd.de](http://www.rzpd.de))

Seq primer: T3 ET from Amersham

High quality sequence stop: 484.

Location/Qualifiers

1..561

/organism="Danio rerio"

/db\_xref="taxon:7955"

/clone="IMAGE:3722356"

/clone\_lib="Zebrafish WashU MPIMG EST"

/sex="mixed"

/tissue\_type="26 somite embryos, adult livers, shield

stage embryos"

/lab\_host="XLI-blue MRF"

/note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; 1st

strand cDNA was primed with a Not I oligo(dT)15 primer

[5'PGACTAGTCTAGATCGGAGCGCCCTTTTITTTTITTTT13];

double-stranded cDNA was ligated to Sal I adaptors (BRL),

digested with Not I and cloned into the Not I and Sal I

sites of the pSPORT1 vector (BRL). Library was constructed

by Matthew Clark (Lehrach lab; ICRF, London and Max Planck

Institut fuer Molekulare Genetik, Berlin). cDNAs for EST

analysis were selected following oligonucleotide

hybridization fingerprinting of arrayed clones from

zebrafish late somitogenesis (26 ss), adult liver or

embryonic shield stage (5.6 h) libraries. Fingerprint

data were used to computationally cluster cDNAs, and a

single cDNA from each cluster was chosen for sequencing.

In some cases multiple members of the same cluster were

sequenced to assess clustering parameters or single clones

were sequenced additional times to assess quality

control."

BASE COUNT 189 a 111 c 161 g 100 t

## ORIGIN

Query Match

Best Local Similarity

Matches

19; Conservative

2.3%; Score 19; DB 9; Length 561;

100.0%; Pred. No. 78;

0; Mismatches

0; Indels

0; Gaps

0;



QY 539 CAGCTGAAGGAGTATCTTT 557  
 |||||  
 Db 31 CAGCTGAAGGAGTATCTTT 49

RESULT 37  
 A2873119 562 bp DNA linear GSS 21-FEB-2001  
 LOCUS  
 DEFINITION 2M0186C14R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC2M0186C14 R, DNA sequence.

ACCESSION A2873119  
 VERSION A2873119  
 KEYWORDS GSS.  
 SOURCE AZ873119.1 GI:13080947  
 ORGANISM house mouse.

ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 562)  
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,  
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
 and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts

JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA

Tel: 801 585 5606  
 Fax: 801 585 7177

Email: dunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0186 Row: C Column: 14

Seq primer: CACACGAGAAACAGTATGACC

Class: plasmid ends

High quality sequence stop: 562.

Location/Qualifiers

1. 562  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC2M0186C14"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42hv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnates/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of PWD42 (gi|4732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."  
 BASE COUNT 144 a 101 c 89 g 228 t  
 ORIGIN

Query Match 2.3%; Score 19; DB 17; Length 562;  
 Best Local Similarity 100.0%; Pred. No. 78;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 593 TTATCACTATTTTAAAGA 611  
 |||||  
 Db 335 TTATCACTATTTTAAAGA 353

RESULT 38  
 A0948387/c

LOCUS  
 DEFINITION Sheared DNA-39J14.TR Sheared DNA Trypanosoma brucei genomic clone  
 Sheared DNA-39J14, DNA sequence.

ACCESSION A0948387  
 VERSION A0948387.1 GI:6771740  
 KEYWORDS GSS.  
 SOURCE Trypanosoma brucei.  
 ORGANISM Trypanosoma brucei

ORGANISM Trypanosoma brucei  
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
 Trypanosoma.

REFERENCE 1 (bases 1 to 608)

AUTHORS El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,  
 Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,  
 Fraser,C. and Adams,M.

TITLE Determination of clone end sequences from Trypanosoma brucei GUTat  
 10.1 sheared DNA library

JOURNAL Unpublished (1999)

COMMENT Other GSSs: Sheared DNA-39J14.TF

Contact: Najib M. El-Sayed

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: nelsayed@tigr.org

Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared  
 DNA library constructed at TIGR. Clones will be available for  
 distribution through ATCC. Sheared DNA end sequences search page:  
<http://www.tigr.org/tdb/mdb/tbdb/>.

Seq primer: M13-Reverse

Class: shotgun.

Location/Qualifiers

1. 608  
 /organism="Trypanosoma brucei"  
 /strain="TREU927/4 GUTat 10.1"  
 /db\_xref="taxon:5691"  
 /clone="Sheared DNA-39J14"  
 /clone\_lib="Sheared DNA-39J14"  
 /note="Vector: pUC18; Site 1: SmaI; Constructed at The  
 Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from a cloned population of  
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically  
 sheared to give a tight size distribution (approx 2 kb).  
 The v + i method used for the library construction is  
 described in detail in Smith, H.O. and Venter, J.C.  
 (Making small insert libraries for whole genome shotgun  
 sequencing projects. In Genome Sequencing: A Practical  
 Approach, eds. M. Vaudin and B. Borell, Oxford University  
 Press, 1999)."

BASE COUNT 114 a 120 c 133 g 241 t

ORIGIN

Query Match 2.3%; Score 19; DB 17; Length 608;

Best Local Similarity 100.0%; Pred. No. 78;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 AATGGAACAAATCTCTCA 205

|||||

Db 105 AATGGAACAAATCTCTCA 87

RESULT 39

A0650378 617 bp mRNA linear EST 18-MAY-2001  
 LOCUS  
 DEFINITION EST328832 tomato germinating seedlings, TAMU Lycopersicon

esculentum cDNA clone cLEI12J12 5', mRNA sequence.

ACCESSION AW650378  
VERSION AW650378.1 GI:7411616  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
REFERENCE 1 (bases 1 to 617)  
AUTHORS Alcala, J., Vrebalov, J., White, R., van der Hoeven, R. S., Holt, I. E.,  
Liang, F., Hansen, T. S., Craven, M. B., Bowman, C. L., Ronning, C. M.,  
Nierman, W., Fraser, C. M., Giovannoni, J. J., Martin, G. B. and Tanksley,  
S. D.

TITLE Generation of ESTs from germinating tomato seed

JOURNAL Unpublished (2000)  
COMMENT Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.

FEATURES  
source  
1..617  
/organism="Lycopersicon esculentum"  
/cultivar="FA496"  
/db\_xref="taxon:4081"  
/clone="cLEI12J12"  
/clone\_lib="tomato germinating seedlings, TAMU"  
/tissue\_type="whole seedlings"  
/dev\_stage="7 days post imbibition"  
/note="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:  
XhoI; 7 days post imbibition on water-agar. Mixed stage  
whole germinating seedlings from seed coat emergence up  
to two centimeters in seeds not showing obvious signs of  
germination were discarded."  
196 a 129 c 97 g 194 t 1 others

BASE COUNT 196 a 129 c 97 g 194 t  
ORIGIN

Query Match 2.3%; Score 19; DB 10; Length 617;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 CAACAATTATTCACAAA 388  
|||||  
DB 387 CAACAATTATTCACAAA 405

RESULT 40  
AW365977/c 619 bp mRNA linear EST 04-FEB-2000  
LOCUS MRO-HT0080-211299-004-e09 HT0080 Homo sapiens cDNA, mRNA sequence.  
DEFINITION  
ACCESSION AW365977  
VERSION AW365977.1 GI:6870627  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 619)  
AUTHORS HCGP <http://www.ludwig.org.br/ORESTES>.  
TITLE The FAPESP/LICR Human Cancer Genome Project  
JOURNAL Unpublished (1999)  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=MRO08t2-MRO-HT0080-211299-004-e09&t3=1999-12-21&t4=1>)  
Seq primer: puc 18 forward  
High quality sequence start: 24  
High quality sequence stop: 495.

FEATURES  
source  
1..619  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT0080"  
/dev\_stage="Adult"  
/note="Organ: head neck; Vector: puc18; Site 1: SmaI;  
Site 2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (O.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
198 a 125 c 135 g 161 t

BASE COUNT 198 a 125 c 135 g 161 t  
ORIGIN

Query Match 2.3%; Score 19; DB 10; Length 619;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 TCAGCAATATGTAGTACTT 483  
|||||  
DB 557 TCAGCAATATGTAGTACTT 539

Search completed: July 4, 2003, 02:33:57  
Job time : 1149 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 8, 2003, 10:33:56 ; Search time 61.5 Seconds  
(without alignments)  
3640.017 Million cell updates/sec

Title: US-10-062-624-41

Perfect score: 1481

Sequence: 1 atgaattataagaataattctt.....ttggaatgaggttcacattc 840

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delpop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2 1/USPTO.spool/US10062624/runat 08072003 093518 1994/app query.fasta\_1.1031  
-DB=A Geneseq 101002 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOPECL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10062624 @CGN 1 114 @runat 08072003 093518 1994 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A Geneseq 101002:\*  
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1462	98.7	280	23	AAU96116
2	1202.5	81.2	283	20	AAU06944
3	1202.5	81.2	283	23	AAU06106
4	1202.5	81.2	283	23	AAU73413
5	778	52.5	165	20	AAU06970
6	644.5	43.5	281	20	AAU06943
7	644.5	43.5	281	20	AAU96105
8	644.5	43.5	281	23	AAU73418
9	629.5	42.5	276	19	AAU51095
10	629.5	42.5	276	21	AAU36189
11	629.5	42.5	276	22	AAU04199
12	622	42.0	286	19	AAU51092
13	621	41.9	288	20	AAU06959
14	620	41.9	286	20	AAU06946
15	620	41.9	286	21	AAU36186
16	620	41.9	286	22	AAU04196
17	620	41.9	286	23	AAU96108
18	620	41.9	286	23	AAU73415
19	605	40.9	280	23	AAU06948
20	605	40.9	280	20	AAU96110
21	603	40.7	280	19	AAU51094
22	603	40.7	280	21	AAU36188
23	603	40.7	280	22	AAU04198
24	603	40.7	280	23	AAU73417
25	601	40.6	280	20	AAU06962
26	601	40.6	280	21	AAU71479
27	601	40.6	280	23	AAU96102
28	592	40.0	285	20	AAU06957
29	592	40.0	285	23	AAU73408
30	588	39.7	307	20	AAU06961
31	582	39.3	278	21	AAU71477
32	582	39.3	278	23	AAU96100
33	577	39.0	278	19	AAU51093
34	577	39.0	278	20	AAU06947
35	577	39.0	278	21	AAU36187
36	577	39.0	278	22	AAU04197
37	577	39.0	278	23	AAU96109
38	577	39.0	278	23	AAU73416
39	567.5	38.3	256	20	AAU06942
40	566.5	38.3	271	23	AAU73419
41	566	38.2	280	19	AAU51089
42	566	38.2	280	21	AAU36183
43	566	38.2	280	22	AAU04193
44	565	38.1	284	23	AAU96111
45	563	38.0	280	20	AAU06945

# ALIGNMENTS

RESULT 1  
AAU96116  
ID AAU96116 standard; Protein; 280 AA.  
XX  
AC AAU96116;  
XX  
DT 02-JUL-2002 (first entry)  
XX  
DE Ehrlichia canis p28-2.  
XX  
KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;  
KW antibacterial.  
XX  
OS Ehrlichia canis.  
XX  
PN WO200222782-A2.  
XX  
PD 21-MAR-2002.  
XX

PF 12-SEP-2001; 2001WO-US28759.  
 XX 12-SEP-2000; 2000US-0660587.  
 XX (RERE-) RES DEV FOUND.  
 XX Walker DH, Yu X, McBride JW;  
 XX WPI; 2002-351882/38.  
 DR N-PSDB; ABK68876.  
 XX  
 XX New recombinant homologous 28 kilodalton immunodominant protein from  
 PT Ehrlichia canis, useful for treating Ehrlichia canis infections  
 XX  
 PS Claim 16; Figure 14; 106pp; English.  
 XX  
 CC The invention relates to a recombinant homologous 28 kDa immunodominant  
 CC protein, p28, (I), of Ehrlichia canis, (I), a 28-kDa antigen preferably  
 CC dispersed in a pharmaceutically acceptable carrier, is useful for  
 CC inhibiting E. canis infection in a subject. (I) is useful in the  
 CC development of vaccines and serodiagnostics that are particularly  
 CC effective for disease prevention and serodiagnostics. AAU96100-AAU96118  
 CC represent the 28-kDa antigen amino acid sequences of the invention.  
 XX  
 SQ Sequence 280 AA;  
 Alignment Scores:  
 Pred. No.: 6.4e-157 Length: 280  
 Score: 1462.00 Matches: 280  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 99.72% Indels: 0  
 DB: 23 Gaps: 0  
 US-10-062-624-41 (1-840) x AAU96116 (1-280)  
 QY 1 ATGATTATAGAAATCTAGTAGAGCGGTTAATCTCATTATCTCAATCTTACCA 60  
 DB 1 MetAsnTyrLysLeuValArgSerAlaLeuIleSerLeuMetSerIleLeuPro 20  
 QY 61 TATCAGTCTTTGACAGCTCTAGGTTCAAGAACTAATGATAACAAAGAGCTTCTAC 120  
 DB 21 TyrGlnSerPheAlaAspProValGlySerArgThrAsnAspAsnLysGluGlyPheTyr 40  
 QY 121 ATTAGTCAAGATACAATCCAAAGTATATACACTTTAGAAATCTCTGCGAAGAACT 180  
 DB 41 IleSerAlaLysTyrAsnProSerIleSerPheArgLysPheSerAlaGluGluThr 60  
 QY 181 CCTATTATGGCAAAATCTCTCACTAAAGATTTTCGGACTAAAGAAAGATGCGTAT 240  
 DB 61 ProIleAsnGlyThrAsnSerLeuThrLysLysValPheGlyLeuLysAspGlyAsp 80  
 QY 241 ATAACAAAAAGACGATTTTACAAGAGTAGCTCCAGGCAATTTTCAAAATAACTTA 300  
 DB 81 IleThrLysLysAspPheThrArgValAlaProGlyIleAspPheGlnAsnLeu 100  
 QY 301 ATATCAGGATTTTCAGGAAGTATGTTACTTATGACGACCGACCAAGAAATAGAACTGAA 360  
 DB 101 IleSerGlyPheSerGlySerIleGlyTyrSerMetAspGlyProArgIleGluGlu 120  
 QY 361 GCTGCATATCAACAATTTAATCCAAACACCGATACCAATGATGATGATGATGATGAA 420  
 DB 121 AlaAlaTyrGlnPheAsnProLysAsnThrAspAsnAspThrAspAsnGlyGlu 140  
 QY 421 TACTATAACATTTTGCATTTCTCGTAAAGATGCAATGGAAGATCAGCAATATAGTA 480  
 DB 141 TyrTyrLysPheAlaLeuSerArgLysAspAlaMetGluAspGlnTyrValVal 160  
 QY 481 CTTAAAAATAGCGGCAATCTTTTATGTCATTTGATGTTGTTGTTGTTGTTGTTGTTGTT 540  
 DB 161 LeuLysAsnAspGlyIleThrPheMetSerLeuMetValAsnThrCysTyrAspIleThr 180  
 QY 541 GCTGAAGGAGTATCTTTCGTACCATATGATGTCAGGATATAGGAGCAGATCTTATCACT 600

Db 181 AlaGluGlyValSerPheValProTyrAlaCysAlaGlyIleGlyAlaAspLeuIleThr 200  
 QY 601 ATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAATAGTATTAGTACCT 660  
 Db 201 IlePheLysAspLeuAsnLeuLysPheAlaTyrGlnGlyLysIleGlyIleSerTyrPro 220  
 QY 661 ATCACACAGAGTCTCTGCATTTTATGTTGGTGGATACCTACCATGCGGTATTGGTAATAA 720  
 Db 221 IleThrProGluValSerAlaPheIleGlyTyrTyrHisGlyValIleGlyAsnLys 240  
 QY 721 TTTGAGAGATACCTGTAATAAATCTCTGTAGTATTAAATGATGCTCTCAAAACACATCT 780  
 Db 241 PheGluLysIleProValIleThrProValValLeuAsnAspAlaProGlnThrThrSer 260  
 QY 781 GCTTCAGTAACTCTTGACGTTGGATCTTTGGCGGAGAAATTTGGAATGAGGTTACCTTC 840  
 Db 261 AlaSerValThrLeuAspValGlyTyrPheGlyGlyGluIleGlyMetArgPheThrPhe 280  
 RESULT 2  
 AAU06944  
 ID AAU06944 standard; Protein; 283 AA.  
 XX  
 AC AAU06944;  
 XX  
 DT 05-JUL-1999 (first entry)  
 XX  
 DE E. chaffeensis OMP-1B protein.  
 XX  
 KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 KW detection; dog.  
 XX  
 OS Ehrlichia chaffeensis.  
 XX  
 FN WO9913720-A1.  
 XX  
 PD 25-MAR-1999.  
 XX  
 PF 18-SEP-1998; 98WO-US19600.  
 XX  
 PR 19-SEP-1997; 97US-0059353.  
 XX  
 PA (OHIS ) UNIV OHIO STATE.  
 XX  
 PI Ohashi N, Rikihisa Y;  
 XX  
 DR WPI; 1999-254290/21.  
 DR N-PSDB; AAX34744.  
 XX  
 PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
 PT Ehrlichia canis  
 XX  
 PS Claim 12; Fig 4B; 55pp; English.  
 XX  
 CC The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 CC in AAU06943-958. The E. canis proteins form part of the P30 family and  
 CC consist of proteins shown in AAU06959-970. The proteins and genes are  
 CC used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX  
 SQ Sequence 283 AA;  
 Alignment Scores:  
 Pred. No.: 1.76e-127 Length: 283  
 Score: 1202.50 Matches: 225  
 Percent Similarity: 88.69% Conservative: 26  
 Best Local Similarity: 79.51% Mismatches: 29  
 Query Match: 81.20% Indels: 3  
 DB: 20 Gaps: 1  
 US-10-062-624-41 (1-840) x AAU06944 (1-283)

```
QY 1 ATGAATTATAGAAATTTCTAGTAGAAGCGCGTTAATCTCATTAAATGTCATCTTACCA 60
Db 1 MetAsnTyrLysLysIlePheValSerSerAlaLeuIleSerLeuMetSerIleLeuPro 20
QY 61 TATCAGCTCTTTGAGATCCCTGTTAGTTCAAGA-----ACTAATGATAACAAAGAA 111
Db 21 TyrGlnSerPheAlaAspProValThrSerAsnAspThrGlyIleAsnAspSerArgGlu 40
QY 112 GGCTTCTACATTAGTGCACAAAGTACAACTCAATCAAGTATATCACATTTAGAAAATTTCTGCT 171
Db 41 GlyPheTyrIleSerValLysTyrAsnProSerIleSerHisPheArgLysPheSerAla 60
QY 172 GAAGAACTCTATTAATGGAACAAATTTCTCTCAATAAAGTTTCGGACTAAAGAAA 231
Db 61 GluGluAlaProIleAsnGlyAsnThrSerIleThrLysLysValPheGlyLeuLysLys 80
QY 232 GATGGTGATATAACAAAAGACGATTTTACAAGAGTAGCTCCAGGCAATTTGATTTCAA 291
Db 81 AspGlyAspIleAlaGlnSerAlaAsnPheAsnArgThrAspProAlaLeuGluPheGln 100
QY 292 AATAACTTAATATCAGGATTTTTCAGGAAGTATTGTTACTCTATGACGGACCAAGAATA 351
Db 101 AsnAsnLeuIleSerGlyPheSerGlySerIleGlyTyrAlaMetAspGlyProArgIle 120
QY 352 GAATCTGAGCTGATATCAACATTTAATCCAAAACACCGATACAAATGATGACTGAT 411
Db 121 GluLeuGluAlaIleTyrGlnLysPheAspAlaLysAsnProAspAsnAspThrAsn 140
QY 412 AATGGTGAATATAACATTTTGCATTTCTCGTAAAGATGCAATGGAAGATCAGCAA 471
Db 141 SerGlyAspTyrTyrLysTyrPheGlyLeuSerArgGluAspAlaIleAlaAspLysLys 160
QY 472 TATGATGACTTAAAAATGACGGCAATACCTTTTATGTCATTGATGGTTAATGCTAT 531
Db 161 TyrValValLeuLysAsnGluGlyIleThrPheMetSerLeuMetValAsnThrCysTyr 180
QY 532 GACATTACAGCTGAAGAGATCTTTCTGACCATATGATGATGTCAGGTATAGGACAGAT 591
Db 181 AspIleThrAlaGluGlyValProPheIleProTyrAlaCysAlaGlyValGlyAlaAsp 200
QY 592 CTATATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAATAGTATT 651
Db 201 LeuIleAsnValPheLysAspPheAsnLeuLysPheSerTyrGlnGlyLysIleGlyIle 220
QY 652 AGTTACCTATCACACAGAGTCTCTGCATTTATTTGGTGATACTACCATGGGTTATT 711
Db 221 SerTyrProIleThrProGluValSerAlaPheIleGlyTyrTyrHisGlyValIle 240
QY 712 GGTAAATTTGAGAAGATACCTGTAATACTCTCTGATGATTAATGATGCTCTCTCAA 771
Db 241 GlyAsnAsnPheAsnLysIleProValIleThrProValValLeuGluGlyAlaProGln 260
QY 772 ACCACATCTGCTTCAGTAACCTTTCAGCTGGATGATCTTTCGCGAGAAATTTGGAATGAG 831
Db 261 ThrThrSerAlaLeuValThrIleAspThrGlyTyrPheGlyGlyGluValGlyValArg 280
QY 832 TTCACCTTC 840
Db 281 PheThrPhe 283
RESULT 3
AAU96106
ID AAU96106 standard; Protein; 283 AA.
XX AC AAU96106;
XX AC AAU96106;
XX 02-JUL-2002 (first entry)
XX DE Ehrlichia chafeensis OMP-1B.
XX KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;
XX KW antibacterial.
```

```
OS Ehrlichia chafeensis.
XX W0200222782-A2.
XX 21-MAR-2002.
XX 12-SEP-2001; 2001WO-US28759.
XX 12-SEP-2000; 2000US-0660587.
XX (RERE-) RES DEV FOUND.
XX Walker DH, Yu X, McBride JW;
XX WPI; 2002-351882/38.
XX New recombinant homologous 28 kilodalton immunodominant protein from
XX Ehrlichia canis, useful for treating Ehrlichia canis infections -
XX Example 3; Figure 3; 106pp; English.
XX The invention relates to a recombinant homologous 28 kDa immunodominant
XX protein, p28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
XX dispersed in a pharmaceutically acceptable carrier, is useful for
XX inhibiting E. canis infection in a subject. (I) is useful in the
XX development of vaccines and serodiagnostics that are particularly
XX effective for disease prevention and serodiagnosis. AAU96100-AAU96118
XX represent the 28-kDa antigen amino acid sequences of the invention.
XX Sequence 283 AA;
Alignment Scores:
Pred. No.: 1,76e-127 Length: 283
Score: 1202.50 Matches: 225
Percent Similarity: 88.69% Conservative: 26
Best-Local Similarity: 79.51% Mismatches: 29
Query Match: 81.20% Indels: 3
DB: 23 Gaps: 1
US-10-062-624-41 (1-840) x AAU96106 (1-283)
QY 1 ATGAATTATAGAAATTTCTAGTAGAAGCGCGTTAATCTCATTAAATGTCATCTTACCA 60
Db 1 MetAsnTyrLysLysIlePheValSerSerAlaLeuIleSerLeuMetSerIleLeuPro 20
QY 61 TATCAGCTCTTTGAGATCCCTGTTAGTTCAAGA-----ACTAATGATAACAAAGAA 111
Db 21 TyrGlnSerPheAlaAspProValThrSerAsnAspThrGlyIleAsnAspSerArgGlu 40
QY 112 GGCTTCTACATTAGTGCACAAAGTACAACTCAATCAAGTATATCACATTTAGAAAATTTCTGCT 171
Db 41 GlyPheTyrIleSerValLysTyrAsnProSerIleSerHisPheArgLysPheSerAla 60
QY 172 GAAGAACTCTATTAATGGAACAAATTTCTCTCAATAAAGTTTCGGACTAAAGAAA 231
Db 61 GluGluAlaProIleAsnGlyAsnThrSerIleThrLysLysValPheGlyLeuLysLys 80
QY 232 GATGGTGATATAACAAAAGACGATTTTACAAGAGTAGCTCCAGGCAATTTGATTTCAA 291
Db 81 AspGlyAspIleAlaGlnSerAlaAsnPheAsnArgThrAspProAlaLeuGluPheGln 100
QY 292 AATAACTTAATATCAGGATTTTTCAGGAAGTATTGTTACTCTATGACGGACCAAGAATA 351
Db 101 AsnAsnLeuIleSerGlyPheSerGlySerIleGlyTyrAlaMetAspGlyProArgIle 120
QY 352 GAATCTGAGCTGATATCAACATTTAATCCAAAACACCGATACAAATGATGACTGAT 411
Db 121 GluLeuGluAlaIleTyrGlnLysPheAspAlaLysAsnProAspAsnAspThrAsn 140
QY 412 AATGGTGAATATAACATTTTGCATTTCTCGTAAAGATGCAATGGAAGATCAGCAA 471
Db 141 SerGlyAspTyrTyrLysTyrPheGlyLeuSerArgGluAspAlaIleAlaAspLysLys 160
```

QY 472 TATGTACTTAAATGACGGCATACCTTTATGTCAATGATGGTTAACTGCTAT 531  
 Db 161 TyrValValLeuLysAsnGluGlyIleThrPheMetSerLeuMetValAsnThrCysTyr 180  
 QY 532 GACATTACAGCTGAAGAGTAGTCTTTCATCATATGATGTCAGAGGTATAGGAGCAGAT 591  
 Db 181 AspileThrAlaGluGlyValProPheIleProTyrAlaCysAlaGlyValGlyAlaAsp 200  
 QY 592 CTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAATAGGTATT 651  
 Db 201 LeuileAsnValPheLysAspPheAsnLeuLysPheSerTyrGlnGlyLysIleGlyIle 220  
 QY 652 AGTTACCTATACACACAGAGTCTCTGATTTATGTTGATGATGATGATGATGATGAT 711  
 Db 221 SerTyrProIleThrProGluValSerAlaPheIleGlyIleGlyTyrTyrHisGlyValIle 240  
 QY 712 GGTAAATAATTTGAGAAGATACCTGTAATACTCTGTAGTATTAAATGATGCTCTCAA 771  
 Db 241 GlyAsnAsnPheAsnLysIleProValIleThrProValValLeuGluGlyAlaProGln 260  
 QY 772 ACCACATCTGCTCAGTAACCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCT 831  
 Db 261 ThrThrSerAlaLeuValThrIleAspThrGlyTyrPheGlyGlyGluValGlyValArg 280  
 QY 832 TTCACCTTC 840  
 Db 281 PheThrPhe 283

## RESULT 4

AAU73413  
 ID AAU73413 standard; Protein; 283 AA.

AC AAU73413;

XX  
 DT 12-MAR-2002 (first entry)

DE Ehrlichia chaffeensis outer membrane protein P28-14.

XX Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.

XX Ehrlichia chaffeensis.

OS  
 PN WO200183699-A2.

XX PD 08-NOV-2001.

XX PF 01-MAY-2001; 2001WO-US13997.

XX PR 01-MAY-2000; 2000US-201035P.

XX (RERE-) RES DEV FOUND.

PA Walker DH, Yu X;

XX WPI; 2002-066527/09.

XX Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated  
 PT P28 useful as a vaccine against Ehrlichia chaffeensis -

XX Disclosure; Figure 2; 97pp; English.

XX The invention relates to isolated and purified 28-kDa outer membrane  
 CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins  
 CC are encoded by a 28kDa outer membrane protein multigene family. P28  
 CC proteins are useful as a vaccine against E. chaffeensis. DNA encoding P28  
 CC is useful for transfecting a host cell. AAU73400-AAU73420 represent  
 CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.

SQ Sequence 283 AA;

Alignment Scores:

Pred. No.: 1.76e-127 Length: 283  
 Score: 1202.50 Matches: 225

Percent Similarity: 88.69% Conservative: 26  
 Best Local Similarity: 79.51% Mismatches: 29  
 Query Match: 81.20% Indels: 3  
 DB: 23 Gaps: 1

US-10-062-624-41 (1-840) x AAU73413 (1-283)

QY 1 ATGAATTAAGAAAAATCTAGTAAGAGCGCGTTAATCTCATTAAATGTCATCTTACCA 60  
 Db 1 MetAsnTyrLysLysIlePheValSerSerAlaLeuIleSerLeuMetSerIleLeuPro 20  
 QY 61 TATCAGCTCTTTTGGCAGATCCTGTAGGTTCAAGA-----ACTAATGATAACAAAGAA 111  
 Db 21 TyrGlnSerPheAlaAspProValThrSerAsnAspThrGlyIleAsnAspSerArgGlu 40  
 QY 112 GGCTTCTACATTAGTCCAAAGTACAATCCAAGTATATCACATTTAGAAAATTTCTGCT 171  
 Db 41 GlyPheTyrIleSerValLysTyrAsnProSerIleSerHisPheArgLysPheSerAla 60  
 QY 172 GAAGAACTCCTATTATTAATGNAACAAATCTCTCACTAAAAGTTTTCGGACTTAAGAA 231  
 Db 61 GluGluAlaProIleAsnGlyAsnThrSerIleThrLysLysValPheGlyLeuLysLys 80  
 QY 232 GATGCTGATATAACAAAAAGACGATTTTCAAGAGTAGCTCCAGGCAATGATTTTCAA 291  
 Db 81 AspGlyAspIleAlaGlnSerAlaAsnPheAsnArgThrAspProAlaLeuGluPheGln 100  
 QY 292 AATAACTTAATATCAGGATTTTCAGGAAGTATTTGTTACTTCTATGACGGACCAAGATA 351  
 Db 101 AsnAsnLeuIleSerGlyPheSerGlySerIleGlyTyrAlaMetAspGlyProArgIle 120  
 QY 352 GAACCTGAGCTGCATATCAACATTTTAACTCAAAAACACCGATACACATGATCTGAT 411  
 Db 121 GluLeuGluAlaLafyrGlnLysPheAspAlaLysAsnProAspAsnAsnAspThrAsn 140  
 QY 412 AATGCTGAATACTATAAACAATTTTGCATTTCTCGTAAAGATGCAATGGAAGATCAGCA 471  
 Db 141 SerGlyAspTyrTyrLysTyrPheGlyLeuSerArgGluAspAlaIleAlaAspLysLys 160  
 QY 472 TATGTACTTAAAAATAGCGCATAACTTTTATGTCATTTGATGTTGTTAATGCTGCTAT 531  
 Db 161 TyrValValLeuLysAsnGluGlyIleThrPheMetSerLeuMetValAsnThrCysTyr 180  
 QY 532 GACATTACAGCTGAAGGAGTATCTTTCGTACCATATGCTGTCAGGTATAGGAGCAGAT 591  
 Db 181 AspIleThrAlaGluGlyValProPheIleProTyrAlaCysAlaGlyValGlyAlaAsp 200  
 QY 592 CTTATCACTATTTTAAAGACCTCAATCTAAATTTTGTCTACCAAGGAAAAATAGGTATT 651  
 Db 201 LeuileAsnValPheLysAspPheAsnLeuLysPheSerTyrGlnGlyLysIleGlyIle 220  
 QY 652 AGTTACCTATACACACAGAGTCTCTGCATTTTATGTTGATGATGATGATGATGATGAT 711  
 Db 221 SerTyrProIleThrProGluValSerAlaPheIleGlyIleGlyTyrTyrHisGlyValIle 240  
 QY 712 GGTAAATAATTTGAGAAGATACCTGTAATACTCTGTAGTATTAAATGATGCTCTCAA 771  
 Db 241 GlyAsnAsnPheAsnLysIleProValIleThrProValValLeuGluGlyAlaProGln 260  
 QY 772 ACCACATCTGCTCAGTAACCTTTCAGCTTTCAGCTTTCAGCTTTCAGCTTTCAGCT 831  
 Db 261 ThrThrSerAlaLeuValThrIleAspThrGlyTyrPheGlyGlyGluValGlyValArg 280  
 QY 832 TTCACCTTC 840  
 Db 281 PheThrPhe 283

## RESULT 5

AAU06970

ID AAU06970 standard; Protein; 165 AA.

XX  
 AC AAU06970;

XX







QY 571 TGTGAGGTATAGGAGCAGATCTTATCACTATTATTTAAAGACCTCAATCTAAATTTGCT 630  
 Db 191 CysAlaGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSer 210  
 QY 631 TACCAAGGAAATAAGTATTAGTTACCTATCACACCAAGAGTCTCTGCAATTTATTCGT 690  
 Db 211 TyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSerValPheIleGly 230  
 QY 691 GGATACTACCATGGCGTTATTGGTAAATAATTTGAGAGATACCTGTGTAATACTCTGTA 750  
 Db 231 GlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrIleIleProThr 250  
 QY 751 GTATTAAATGATGCTCTCA---ACCACATCTGCTTCAGTAACCTCTTGAGGTGGATAC 807  
 Db 251 GlySerThrLeuAlaGlyLysGlyAsnTyrProAlaIleValIleLeuAspValCysHis 270  
 QY 808 TTTCGCGAGAAATGGGAATGAGGTTCACCTTC 840  
 Db 271 PheGlyIleGluLeuGlyGlyArgPheAlaPhe 281

## RESULT 8

AAU73418  
 ID AAU73418 standard; Protein; 281 AA.

XX AC AAU73418;

XX DT 12-MAR-2002 (first entry)

XX DE Ehrlichia chaffeensis outer membrane protein P28-19.

XX KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.

XX OS Ehrlichia chaffeensis.

XX XX WO200183699-A2.

XX PD 08-NOV-2001;

XX PF 01-MAY-2001; 2001WO-US13997.

XX PR 01-MAY-2000; 2000US-201035P.

XX PA (RERE-) RES DEV FOUND.

XX PI Walker DH, Yu X;

XX DR WPI; 2002-066527/09.

XX PT Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated P28 useful as a vaccine against Ehrlichia chaffeensis -

XX PS Disclosure; Figure 2; 97pp; English.

XX CC The invention relates to isolated and purified 28-kDa outer membrane proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins are encoded by a 28kDa outer membrane protein multigene family. P28 proteins are useful as a vaccine against E. chaffeensis. DNA encoding P28 is useful for transfecting a host cell. AAU73400-AAU73420 represent Ehrlichia chaffeensis P28 outer membrane proteins of the invention.

XX SQ Sequence 281 AA;

## Alignment Scores:

Pred. No.: 3,51e-64 Length: 281  
 Score: 644.50 Matches: 140  
 Percent Similarity: 62.20% Conservative: 41  
 Best Local Similarity: 48.11% Mismatches: 89  
 Query Match: 43.52% Indels: 21  
 DB: 23 Gaps: 6

US-10-062-624-41 (1-840) x AAU73418 (1-281)

QY 1 ATGAATTATAGAAAATCTAGTAGAAGCGCGTTAATCTCAATTAATGTCATCTTACCA 60

Db 1 MetAsnTyrLysLysValPheIleThrSerAlaLeuIleSerSerLeuPro 20  
 QY 61 TATCAGCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGATAACAAAGAGGCTTCTTAC 120  
 Db 21 GlyValSerPheSerAspProAlaGlySerGlyIleAsnGlyAsn-----PheTyr 37  
 QY 121 ATTAGTCAAAAGTACAATCCAAAGTATATCACACTTTTGAATAATTTCTCTGTGAAAGAACT 180  
 Db 38 IleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu--- 56  
 QY 181 CCTATTAAATGGAACAAATTTCTCTCACTAAATAAGTTTTTCGACTAAAGAAAGATGGTGAT 240  
 Db 57 -----GluArgAsnThrThrValGlyValPheGlyLeuLysGlnAsnTrpAsp 72  
 QY 241 -----ATAACAAAAGAACCATTTTACAAGAGTAGTCCAGGCATT 282  
 Db 73 GlySerAlaIleSerAsnSerSerProAsnAspValPheThrValSerAsnTyrSerPhe 92  
 QY 283 GATTTTCAAAATAACTTAATATCAGAGATTTTTCAGAAAGATTTGTTACTCTATCGACGGA 342  
 Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaIleGlyIleGlyTyrSerMetAspGly 112  
 QY 343 CCAAGATAGAACTTGAAGCTGCATATCAACAATTTTAATCCAAAACACCGATAACAAT 402  
 Db 113 ProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn 132  
 QY 403 GATACTGATAATGTTGAATCTATATAACATTTTTCATTTCTCGTAAAGATGCAATGAA 462  
 Db 133 -----TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerHisAsnSerAlaAlaAsp 150  
 QY 463 -----GATCAGCAATATGTAGTACTTAAATAATGACGGCATAACTTTTATGTCA 510  
 Db 151 MetSerSerAlaSerAsnAsnPheValPheLeuLysAsnGluGlyLeuLeuAspIleSer 170  
 QY 511 TTGATGTTAATCTTCTATGATCATTACAGCTGAAGAGATATCTTTCGTACCATATGCA 570  
 Db 171 PheMetLeuAsnAlaCysTyrAspValValGlyGlyIleProPheSerProTyrIle 190  
 QY 571 TGTGAGGTATAGGAGCAGATCTTATCACTATTTTTAAAGACCTCAATCTAAATTTGCT 630  
 Db 191 CysAlaGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSer 210  
 QY 631 TACCAAGGAAATAAGTATTAGTTACCTATCACACCAAGAGTCTCTGCAATTTATTCGT 690  
 Db 211 TyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSerValPheIleGly 230  
 QY 691 GGATACTACCATGGCGTTATTGGTAAATAATTTGAGAGATACCTGTGTAATACTCTGTA 750  
 Db 231 GlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrIleIleProThr 250  
 QY 751 GTATTAAATGATGCTCTCA---ACCACATCTGCTTCAGTAACCTCTTGAGGTGGATAC 807  
 Db 251 GlySerThrLeuAlaGlyLysGlyAsnTyrProAlaIleValIleLeuAspValCysHis 270  
 QY 808 TTTCGCGAGAAATGGGAATGAGGTTCACCTTC 840  
 Db 271 PheGlyIleGluLeuGlyGlyArgPheAlaPhe 281

## RESULT 9

AAW51095

ID AAW51095 standard; Protein; 276 AA.

XX AC AAW51095;

XX DT 14-SEP-1998 (first entry)

XX DE Ehrlichia chaffeensis VSA5 protein (partial sequence).

XX KW MAP1 homologue; variable surface antigen; VSA5; rickettsia; DNA vaccine.

XX OS Ehrlichia chaffeensis.



DR N-PSDB; AAC68706.

XX New polynucleotides useful as DNA vaccines for conferring immunity to  
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,  
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens  
 PT  
 XX

PS Claim 3; Page 47; 63pp; English.

XX The present sequence shows a high degree of similarity to the major  
 CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be  
 CC used in a vaccines to protect animals or humans against rickettsial  
 CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,  
 CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response  
 CC protective against the rickettsial pathogen. The nucleic acid vaccines  
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.  
 CC Cowdria ruminantium genes designated map 2, lhwr3f, 4hwr3f, 18hwr3f  
 CC and 3gdr3f may be used in therapeutic and diagnostic applications. The  
 CC polypeptides are useful for detecting antibodies associated with  
 CC infection by a rickettsial pathogen whilst the polynucleotides may be  
 CC used to detect the presence of rickettsial nucleic acids.

XX Sequence 276 AA;

## Alignment Scores:

Pred. No.: 1.75e-62 Length: 276  
 Score: 629.50 Matches: 137  
 Percent Similarity: 62.24% Conservative: 41  
 Best Local Similarity: 47.90% Mismatches: 87  
 Query Match: 42.51% Indels: 21  
 DB: 21 Gaps: 6

US-10-062-624-41 (1-840) x AAB36189 (1-276)

QY 1 ATGAATTATAAGAAATCTCTAGAGAGCGCTTAATCTCATTAATGCTCAATCTTACCA 60  
 DB 1 MetAsnTyrLysLysValPheIleThrSerAlaLeuIleSerLeuIleSerLeuPro 20  
 QY 61 TATCAGTCTTTTGCAGATCTGTAGGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120  
 DB 21 GlyValSerPheSerAspProAlaGlySerGlyIleAsnGlyAsn-----PheTyr 37  
 QY 121 ATTAGTCAAGTACAAATCCAGTATATACACATTTAGAAAATTTCTCTGTGAAGAACT 180  
 DB 38 IleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu--- 56  
 QY 181 CCTATTATGCAACAAATCTCTCACTAAAGTTTTCGACTAAAGAGAGATGGTCAAT 240  
 DB 57 -----GluArgAsnThrThrValGlyValPheGlyLeuLysGlnAsnTrpAsp 72  
 QY 241 -----ATAACAAAAAGACGATTTTACAAGAGTAGCTCCAGGCATT 282  
 DB 73 GlySerAlaIleSerAsnSerSerProAsnAspValPheThrValSerAsnTyrSerPhe 92  
 QY 283 GATTTTCAAAATACTTAATATCAGGATTTTCAGAGATATGTTACTCTATCGACGGA 342  
 DB 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetAspGly 112  
 QY 343 CCAAGATAGCACTTGAAGCTGCATATCAACAATTTTAAATCCAAAAACCCGATACAACT 402  
 DB 113 ProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsn 132  
 QY 403 GATACTGATAATGGTGAATACATAAATTTTGCATTATCTCGTAAAGATGCAATGAA 462  
 DB 133 -----TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerHisAsnSerAlaAsp 150  
 QY 463 -----GATCAGCAATATGTACTATTAAAAATGACGCAATACTTTTATGTCA 510  
 DB 151 MetSerSerAlaSerAsnAsnPheValPheLeuLysAsnGluGlyLeuLeuAspIleSer 170  
 QY 511 TTGATGGTTATCTTGTATGACATACAGCTCAAGAGATATCTTTCGATACCATATGCA 570  
 DB 171 PheMetLeuAsnAlaCysTyrAspValValGlyGluGlyIleProPheSerProTyrIle 190

QY 571 TGTCCAGGTATAGCAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCT 630  
 DB 191 CysAlaGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSer 210  
 QY 631 TACCAAGCAAAATAGGTATTTAGTACCTATACACACAGAGTCTCTGCATTTTGGT 690  
 DB 211 TyrGlnGlyLysLeuGlyLysSerTyrSerIleSerProGluAlaSerValPheIleGly 230  
 QY 691 GGATACTACCATGGCGTTATTGGTAATAAATTTTGAAGATACCTGTAAATACCTCTGTA 750  
 DB 231 GlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrIleIleProThr 250  
 QY 751 GTATTAAATGATCTCTCTCAA---ACCACATCTCTCAGTAACCTTTGAGTTGGATAC 807  
 DB 251 GlySerThrLeuAlaGlyLysGlyAsnTyrProAlaIleValIleLeuAspValCysHis 270  
 QY 808 TTTGGCGGAGAAATTTGGA 825  
 DB 271 PheGlyIleGluMetGly 276

## RESULT 11

AAU04199  
 ID AAU04199 standard; Protein; 276 AA.

XX AC AAU04199;

XX DT 23-OCT-2001 (first entry)

XX DE Variable surface antigen 5 (VSA5) from Ehrlichia chaffeensis.

XX KW Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;  
 XX infection; heartwater; diagnostic; variable surface antigen; VSA.

XX OS Ehrlichia chaffeensis.

XX FN US6251872-B1.

XX XX 26-JUN-2001.

XX PF 17-OCT-1997; 97US-0953326.

XX PR 17-OCT-1996; 96US-0733230.

XX XX (UYFL) UNIV FLORIDA.

XX PI Barbet AF, Ganta RR, McGuire TC, Burrige MJ, Nyika A;  
 XX PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;

XX DR WPI; 2001-424487/45.

XX DR N-PSDB; AAS07578.

XX PT New MAP2 genes and polypeptides useful as vaccines for conferring  
 PT immunity to human and animal rickettsial diseases, e.g. heartwater, or  
 PT as molecular markers in nucleic acid analysis procedures -

XX PS Example 3; Fig 2A-2B; 30pp; English.

XX CC The sequence represents the amino acid sequence of variable surface  
 CC antigen 5 (VSA5) isolated from Ehrlichia chaffeensis, which  
 CC has similarity to major antigen protein (MAP). The MAP polynucleotides  
 CC and polypeptides are useful as vaccines for conferring immunity to  
 CC rickettsia infection, including Cowdria ruminantium causing heartwater.  
 CC The MAP polynucleotides may be used as molecular markers in nucleic acid  
 CC analysis procedures, and to produce the MAP polypeptides, which may  
 CC be used to raise antibodies that are reactive with the polypeptides.  
 CC The nucleic acids may further be used as probes to identify  
 CC complementary sequences within other nucleic acid molecules or genomes,  
 CC where such probes can be applied to identify or distinguish infectious  
 CC strains of organisms in diagnostic procedures or in rickettsial  
 CC research where identification of particular organisms or strains is  
 CC needed.

XX CC





Db 171 AsnGluGlyLeuLeuAspIleSerLeuMetIleAsnAlaCysTyrAspIleThrIleAsp 190  
 Qy 547 GGAGTATCTTCGTACCATATGATGTCAGGTATAGGACGAGATCTTATCACTATTTT 606  
 Db 191 SerMetProPheSerProTyrIleCysAlaGlyIleGlySerAspLeuValSerMetPhe 210  
 Qy 607 AAGACCTCAATCTAAATTTGCTTACCAGGAGAAATAGGTATTAGTACCCTATCACA 666  
 Db 211 GluThrThrAsnProLysIleSerTyrGlnGlyLysLeuGlyValSerTyrSerIleSer 230  
 Qy 667 CCAGAAAGTCTCTGCATTTATTGGTGGATACCATGGCTTATTGGTAAATAATTTGAG 726  
 Db 231 ProGluAlaSerValPheValGlyHisPheHisArgValIleGlyAsnGluPheLys 250  
 Qy 727 AAGATACCTGTAATCACTCTGCTAGTATTAAATGATGCTCCCAACCAACACATCTGCTTCA 786  
 Db 251 AspIleProAlaIleThrProAlaGlyAlaThrGluIleLysGlyThrGlnPheThrThr 270  
 Qy 787 GTAACCTTGAGTTGGATCTTTCGGCGAGAAATTTGGAATGAGCTTCACCTTC 840  
 Db 271 ValThrLeuAsnIleCysHisPheGlyLeuGluLeuGlyGlyArgPheThrPhe 288

## RESULT 14

ID AAY06946 standard; Protein; 286 AA.  
 AC AAY06946;

DT 05-JUL-1999 (first entry)  
 DE E. chaffeensis OMP-1D protein.

KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 KW detection; dog.  
 OS Ehrlichia chaffeensis.  
 FN WO9913720-A1.  
 PD 25-MAR-1999.  
 PF 18-SEP-1998; 98WO-US19600.  
 PR 19-SEP-1997; 97US-0059353.  
 XX (OHIS ) UNIV OHIO STATE.  
 PI Ohashi N, Rikihisa Y;  
 DR WPI: 1999-254290/21.  
 DR N-PSDB; AAX34746.

PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
 PT Ehrlichia canis  
 XX Claim 14; Fig 6B; 55pp; English.

CC The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and  
 CC consist of proteins shown in AAY06959-970. The proteins and genes are  
 CC used to detect E. chaffeensis in patients and E. canis in dogs.

XX Sequence 286 AA;

## Alignment Scores:

Pred. No.:	2,12e-61	Length:	286
Score:	620.00	Matches:	139
Percent Similarity:	61.56%	Conservative:	42
Best Local Similarity:	47.28%	Mismatches:	89
Query Match:	41.86%	Indels:	24
DB:	20	Gaps:	7

US-10-062-624-41 (1-840) x AAY06946 (1-286)  
 Qy 1 ATGAATATAGAAAAATTTCTAGTAAGAAGCGCGTTAATCTCATTAATGTCTCAATCTTACCA 60  
 Db 1 MetAsnGlyLysPhePheIleThrAlaLeuThrLeuLeuMetSerPheLeuPro 20  
 Qy 61 TATCAGTCTTTTGAGATCCCTGTAGGTTCAGAAGTCAAGTAAATGATATACAAAGAGGC---TTC 117  
 Db 21 GlyIleSerLeuSerAspProVal-----GlnAspAspAsnIleSerGlyAsnPhe 37  
 Qy 118 TACATTAGTCAAAAGTACAATCCAAGTATATACACATTTTAGAAAAATCTCTGCTCAAGAA 177  
 Db 38 TyrIleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu 57  
 Qy 178 ACTCCTTAATGAACAAATTTCTCTCACTAAAAAGTTTTCGCACTAAAGAAAGATGGT 237  
 Db 58 -----GluArgAsnThrThrValGlyValPheGlyIleGluGlnAspTyr 72  
 Qy 238 GAT-----ATACAAAAAAGACGATTTTACAAGAGTAGTCCAGGC 279  
 Db 73 AspArgCysValIleSerArgThrLeuSerAspIlePheThrValProAsnTyrSer 92  
 Qy 280 ATTGATTTTCAAAATACTTAATATCAGGATTTTTCAGGAAGTATTGGTTACTTATGGAC 339  
 Db 93 PheLysTyrGluAsnAsnLeuPheSerGlyPheAlaGlyAlaIleGlyTyrSerMetAsp 112  
 Qy 340 GGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGATAAC 399  
 Db 113 GlyProArgIleGluLeuGluValSerTyrGluAlaPheAspValLysAsnGlnGlyAsn 132  
 Qy 400 AATGATCTGATAATGGTGAATACTATATAAATTTTCGATTTATCTCGTAAAGATCCATG 459  
 Db 133 Asn-----TyrLysAsnGluAlaHisArgTyrTyrAlaLeuSerHisLeuLeuGlyThr 150  
 Qy 460 GAAGATCAG-----CAATATCTAGTATTAATAAATGACCGCATAACT 501  
 Db 151 GluThrGlnIleAspGlyAlaGlySerAlaSerValPheLeuIleAsnGluLeuLeu 170  
 Qy 502 TTTATGTCATTGATGGTTAATCTTGTATGATGATACAGTACAGTGAAGGAGTATCTTTGCTA 561  
 Db 171 AspLysSerPheMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSer 190  
 Qy 562 CCATATGTCATGTCAGGTATAGGACAGTCTTATCACTATTATTTTAAAGACTCAATCTA 621  
 Db 191 ProTyrIleCysAlaGlyIleGlyLeuAspLeuValSerMetPheGluAlaIleAsnPro 210  
 Qy 622 AAATTTGCTTACCAGGAAAAATAGTATTAGTTACCTTATCACACCAAGAAAGTCTCTGCA 681  
 Db 211 LysIleSerTyrGlnGlyLysLeuGlyLeuSerTyrProIleSerProGluAlaSerVal 230  
 Qy 682 TTTATGTCGATACTACCATGGCGTTATTGGTAAATAATTTGAGAAGATACCTGTAATA 741  
 Db 231 PheIleGlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrMet 250  
 Qy 742 ACTCTGTAGTATTAAATGATGCTCTCAAA---ACCACATCTGCTTCACTAGTAACTTCGAC 798  
 Db 251 IleProSerGluSerAlaLeuAlaGlyLysGlyLysGlySerTyrProAlaIleValThrLeuAsp 270  
 Qy 799 GTTGATACTTTGGCGGAGAAATTTGGAATGAGGTTCACCTTC 840  
 Db 271 ValPheTyrPheGlyIleGluLeuGlyGlyArgPheAsnPhe 284

## RESULT 15

ID AAB36186  
 AC AAB36186;

DT 02-MAR-2001 (first entry)  
 DE Ehrlichia chaffeensis partial VSA2.  
 XX



PT New MAP2 genes and polypeptides useful as vaccines for conferring  
 PT immunity to human and animal rickettsial diseases, e.g. heartwater, or  
 PT as molecular markers in nucleic acid analysis procedures -  
 XX  
 PS  
 PS Example 3; Fig 2A-2B; 30pp; English.

CC The sequence represents the amino acid sequence of variable surface  
 CC antigen 2 (VSA2) isolated from Ehrlichia chaffeensis, which  
 CC has similarity to major antigen protein (MAP). The MAP polynucleotides  
 CC and polypeptides are useful as vaccines for conferring immunity to  
 CC rickettsia infection, including Cowdria ruminantium causing heartwater.  
 CC The MAP polynucleotides may be used as molecular markers in nucleic acid  
 CC analysis procedures, and to produce the MAP polypeptides, which may  
 CC be used to raise antibodies that are reactive with the polypeptides.  
 CC The nucleic acids may further be used as probes to identify  
 CC complementary sequences within other nucleic acid molecules or genomes,  
 CC where such probes can be applied to identify or distinguish infectious  
 CC strains of organisms in diagnostic procedures or in rickettsial  
 CC research where identification of particular organisms or strains is  
 CC needed.

SQ Sequence 286 AA;

Alignment Scores:  
 Pred. No.: 2,12e-61 Length: 286  
 Score: 620.00 Matches: 139  
 Percent Similarity: 61.56% Conservative: 42  
 Best Local Similarity: 47.28% Mismatches: 89  
 Query Match: 41.86% Indels: 24  
 DB: 22 Gaps: 7

US-10-062-624-41 (1-840) x AAU04196 (1-286)

QY 1 ATGAATTATAAGAAATCTAGTAAAGCGGTAACTCTAATTAATGTCATCTTACCA 60  
 Db 1 MetAsnCysGluLysPhePheileThrThrAlaLeuThrLeuLeuMetSerPheLeuPro 20  
 QY 61 TATCAGTCTTTGCGAGCTCTAGGTTCAGAACTAATGATAACAAAGAGGC---TTC 117  
 Db 21 GlyIleSerLeuSerAspProVal-----GlnAspAsnIleSerGlyAsnPhe 37  
 QY 118 TACATTAGTCAAGTACAAATCCAAATATATACACATTTAGAAAATCTCTCTCAAGAA 177  
 Db 38 TyrIleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu 57  
 QY 178 ACTCTATTATGGAACAATAATCTCTCACTAAAGAGTTTCGGACTAAAGAAAGAGTGT 237  
 Db 58 -----GluArgAsnThrThrValGlyValPheGlyIleGluGlnAspTrp 72  
 QY 238 GAT-----ATAACAAAAAAGAGCATTTTACAAGAGTAGTCCAGGC 279  
 Db 73 AspArgCysValIleSerArgThrThrLeuSerAspIlePheThrValProAsnTyrSer 92  
 QY 280 ATTGATTTTCAAAATAACTTAATATCAGGATTTTCAAGAGTATTGGTCTATCGAC 339  
 Db 93 PheLysTyrGluAsnAsnLeuPheSerGlyPheAlaGlyAlaIleGlyTyrSerMetAsp 112  
 QY 340 GGACCAAGATGAAGCTGATATCAACAATTTAATCCAAAAAACCCGATAAC 399  
 Db 113 GlyProArgIleGluLeuGluValSerTyrGluAlaPheAspValIleAsnGlnGlyAsn 132  
 QY 400 RATGATCTGATAATGGTGAATCTATATAACATTTTGCATTATCTCGTAAAGATCAATG 459  
 Db 133 Asn-----TyrLysAsnGluAlaHisArgTyrTyrAlaLeuSerHisLeuLeuGlyThr 150  
 QY 460 GAAGATCAG-----CAATATGTAGTACTTTAAAAATGACCGCATAACT 501  
 Db 151 GluThrGlnIleAspGlyAlaGlySerAlaSerValPheLeuIleAsnGluGlyLeuLeu 170  
 QY 502 TTTATGCTATGAGTGAATTAATCTGCTATGACATTTACAGCTGAAGGAGTATCTTCGTA 561  
 Db 171 AspLysSerPheMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSer 190

QY 562 CCATATGATCATGTGAGGTATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTA 621  
 Db 191 ProTyrIleCysAlaGlyIleGlyLeuAspLeuValSerMetPheGluAlaIleAsnPro 210  
 QY 622 AAATTTGCTTACCAAGGAAAAATAGTATTAGTTTACCTATCACACCAAGAGTCTCTCCA 681  
 Db 211 LysIleSerTyrGlnGlyLysLeuGlyLeuSerTyrProIleSerProGluAlaSerVal 230  
 QY 682 TTTATGCTGATACTACCATGGGCTTATGCTTAATAAATTTGAGAAGATACCTGTAATA 741  
 Db 231 PheIleGlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrMet 250  
 QY 742 ACTCTGTAGTATTAAATGATGCTCCTCAA---ACCACATCTGCTTCACTAACTCTTCAC 798  
 Db 251 IleProSerGluSerAlaLeuAlaGlyLysGlyAsnTyrProAlaIleValThrLeuAsp 270  
 QY 799 GTTGGATCTTTGGCGGAGAAATTGGAATGAGGTTTCACCTTC 840  
 Db 271 ValPheTyrPheGlyIleGluLeuGlyGlyArgPheAsnPhe 284

RESULT 17

AAU96108  
 ID AAU96108 standard; Protein; 286 AA.

AC AAU96108;  
 XX  
 DT 02-JUL-2002 (first entry)  
 XX  
 DE Ehrlichia chaffeensis OMP-1D.  
 XX  
 KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;  
 XX  
 OS Ehrlichia chaffeensis.  
 XX  
 PN WO200222782-A2.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PF 12-SEP-2001; 2001WO-US28759.  
 XX  
 PR 12-SEP-2000; 2000US-0660587.  
 XX  
 PA (RESE-) RES DEV FOUND.  
 XX  
 PI Walker DH, Yu X, McBride JW.  
 XX  
 DR WPI; 2002-351882/38.  
 XX  
 PT New recombinant homologous 28 kilodalton immunodominant protein from  
 PT Ehrlichia canis, useful for treating Ehrlichia canis infections -  
 XX

Example 3; Figure 3; 106pp; English.

CC The invention relates to a recombinant homologous 28 kDa immunodominant  
 CC protein, p28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably  
 CC dispersed in a pharmaceutically acceptable carrier, is useful for  
 CC inhibiting E. canis infection in a subject. (I) is useful in the  
 CC development of vaccines and serodiagnostics that are particularly  
 CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118  
 CC represent the 28-kDa antigen amino acid sequences of the invention.

SQ Sequence 286 AA;

Alignment Scores:  
 Pred. No.: 2,12e-61 Length: 286  
 Score: 620.00 Matches: 139  
 Percent Similarity: 61.56% Conservative: 42  
 Best Local Similarity: 47.28% Mismatches: 89  
 Query Match: 41.86% Indels: 24  
 DB: 23 Gaps: 7

US-10-062-624-41 (1-840) x AAU96108 (1-286)



Qy	1	ATGAATTATAGAA	AAATCTCTAGTAAGACGCGGTTAATCTCAATTAATGTCAATCTTACCA	60
Db	1	MetAenCysGluLysPhe	PheIleThrAlaLeuThrLeuLeuMetSerPheLeuPro	20
Qy	61	TATCAGTCTTTTG	CAGATCCCTGTAGGTTCAAGAACTAATGATAACAAAGAACGC--TTC	117
Db	21	GlyIleSerLeuSer	AspProVal-----GlnAspAspAsnIleSerGlyAenPhe	37
Qy	118	TACATTAGTCAA	AGTACAATCCAAAGTATATACACACTTTAGAAAAATCTCTCTCGTCAAGAA	177
Db	38	TyrIleSerGlyLys	TyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu	57
Qy	178	ACTCCTATTATG	NACAAATCTCTCACTAAAAAGTTTCGGACTAAGAAGAAATGGT	237
Db	58	-----GluArg	AsnThrThrValGlyValPheGlyIleGluGlnAspTrp	72
Qy	238	GAT-----ATA	CAAAAAAGACGATTTTACAAAGAGTAGCTCCAGGC	279
Db	73	AspArgCysValI	leSerArgThrThrLeuSerAspIlePheThrValProAsnTyrSer	92
Qy	280	ATTGATTTTCAA	ATAACTTAATACAGATTTTCAGGAAGPATTTGGTTACTCTATGGAC	339
Db	93	PheLysTyrGlu	AsnAsnLeuPheSerGlyPheAlaGlyAlaIleGlyTyrSerMetAsp	112
Qy	340	GGACCAAGATTA	GAAGCTTGAAGTCCATATCAACAATTTAATCCAAAAACACCATATAC	399
Db	113	GlyProArgIle	GluLeuGluValSerTyrGluAlaPheAspValLysAsnGlnGlyAsn	132
Qy	400	AATGATACTGTA	TATGGTGAATCATATAACAATTTTGCAATTTATCTCGTAAAGATGCAATG	459
Db	133	Asn-----Tyr	LysAsnGluAlaHisArgTyrTyrAlaLeuSerHisLeuLeuGlyThr	150
Qy	460	GAAGATCAG-----	CAATATGTAGTACTTTAAAAATGACGGCATAACT	501
Db	151	GluThrGlnIle	AspGlyAlaGlySerAlaSerValPheLeuIleAsnGluGlyLeuLeu	170
Qy	502	TTTATGTCAAT	TGGTTAAATACTTGCTATGACATTACAGCTGAGGAGTAGTCTTTCGTA	561
Db	171	AspLysSerPhe	MetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSer	190
Qy	562	CCATATGCATG	TCAGGTPATAGGACAGACTTTATCACTATTTTAAAAACCTCAATCTA	621
Db	191	ProTyrIleCys	AlaGlyIleGlyIleAspLeuValSerMetPheGluAlaIleAsnPro	210
Qy	622	AAATTTGCTTAC	CAAGGAAAAATAGGTATTAGTTACCTATCACACCAAGACTCTCTGCA	681
Db	211	LysIleSerTyr	GinGlyLysLeuGlyLeuSerTyrProIleSerProGluAlaSerVal	230
Qy	682	TTTATTCGTGT	ACTACTACCATCGGCTTATCGGTAAATAATTGAGAAAGATACCTGTAATA	741
Db	231	PheIleGlyHis	PheHisGlyValIleGlyAsnGluPheArgAspIleProThrMet	250
Qy	742	ACTCCTGTAGT	PATTAATGATGCTCCTCAA---ACCATCATCTGCTTCAGTAAGTCTTGAC	798
Db	251	IleProSerGlu	SerAlaLeuAlaGlyLysGlyAsnTyrProAlaIleValThrLeuAsp	270
Qy	799	GTTGGATACTT	TGGCGGAGAAATTGGAAAGAGGTTTCACCTTC	840
Db	271	ValPheTyrPhe	GlyIleGluLeuGlyGlyArgPheAsnPhe	284

RESULT 18  
AAU73415  
ID AAU73415 standard; Protein; 286 AA.  
XX  
XX  
AC AC  
AAU73415;  
XX  
XX  
DT 12-MAR-2002 (first entry)  
XX  
DE Ehrlichia chaffeensis outer membrane protein P28-16.  
XX  
XX  
KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.  
XX

OS	Ehrlichia chaffeensis.
XX	
PN	WO200183699-A2.
XX	
PD	08-NOV-2001.
XX	
Pf	01-MAY-2001; 2001WO-US13997.
XX	
PR	01-MAY-2000; 2000US-201035P.
XX	
PA	(RERE-) RES DEV FOUND.
XX	
PI	Walker DH, Yu X;
XX	
DR	WPI; 2002-066527/09.
XX	
PT	Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated
PT	P28 useful as a vaccine against Ehrlichia chaffeensis -
XX	
PS	Disclosure; Figure 2; 97pp; English.
XX	
CC	The invention relates to isolated and purified 28-kDa outer membrane
CC	proteins (p28-1 to p28-21) of Ehrlichia chaffeensis. P28 proteins
CC	are encoded by a 28kba outer membrane protein multigene family. P28
CC	proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28
CC	is useful for transfecting a host cell. AAU73400-AAU73420 represent
CC	Ehrlichia chaffeensis P28 outer membrane proteins of the invention.
XX	
SQ	Sequence 286 AA;
 Alignment Scores:	
Pred. No.:	2,12e-61 Length: 286
Score:	620.00 Matches: 139
Percent Similarity:	61.56% Conservative: 42
Best Local Similarity:	47.28% Mismatches: 89
Query Match:	Indels: 24
DB:	Gaps: 7
 US-10-062-624-41 (1-840) x AAU73415 (1-286)	

```

151 GluThrGlnIleAspGlyAlaGlySerAlaSerValPheLeuIleAsnGluGlyLeuLeu 170
152 TTTATGTCATGATGTTAAATCTTGCTATGACATTACAGCTGAAGGAGTATCTTTCTGTA 561
171 AspLysSerPheMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSer 190
562 CCATATGATGTGACGAGTATAGGACAGACTTATACACTATTTTAAAGACCTCAATCTA 621
191 ProTyrIleCysAlaGlyIleGlyIleAspLeuValSerMetPheGluAlaIleAsnPro 210
622 AAATTTGCTTACCAAGGAAATAGTATTAGTTACCTCTATCACACCAAGAGCTCTCTGCA 681
211 LysIleSerTyrGlnGlyLysLeuGlyLeuSerTyrProIleSerProGluAlaSerVal 230
682 TTTATGTTGGTACTACCATGGCTTATGCTTAATAAATTTGAGAAGATACCTGTAATA 741
231 PheIleGlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrMet 250
742 ACTCCTGTAGTATTAAATGATGCTCTCA---ACACATCTGCTTCAGTAACTCTTGAC 798
251 IleProSerGluSerAlaLeuAlaGlyLysGlyAsnTyrProAlaIleValThrLeuAsp 270
799 GTTGATCTTTGGCGGAGAAATGGAATGAGGTTTCACCTTC 840
271 ValPheTyrPheGlyIleGluLeuGlyGlyArgPheAsnPhe 284

```

## RESULT 19

AAU06948  
ID AAY06948 standard; Protein; 280 AA.

AC AAY06948;

DT 05-JUL-1999 (first entry)

DE E. chaffeensis OMP-1F protein.

KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
KW detection; dog.

OS Ehrlichia chaffeensis.

PN W09913720-A1.

PD 25-MAR-1999.

PF 18-SEP-1998; 98WO-US19600.

PR 19-SEP-1997; 97US-0059353.

PA (OHIS ) UNIV OHIO STATE.

PI Ohashi N, Rikihisa Y;

DR WPI: 1999-254290/21.

DR N-PSDB; AAX34748.

XX Novel outer membrane proteins from Ehrlichia chaffeensis and  
PT Ehrlichia canis

PS Claim 16; Fig 8B; 55pp; English.

XX The invention provides isolated outer membrane proteins (OMP) from  
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
CC in AAY06943-958. The E. canis proteins form part of the P30 family and  
CC consist of proteins shown in AAY06959-970. The proteins and genes are  
CC used to detect E. chaffeensis in patients and E. canis in dogs.

SQ Sequence 280 AA;

Alignment Scores:

Pred. No.: 1.06e-59 Length: 280

```

Score: 605.00 Matches: 133
Percent Similarity: 63.39% Conservative: 54
Best Local Similarity: 45.08% Mismatches: 78
Query Match: 40.85% Indels: 30
DB: 20 Gaps: 10

US-10-062-624-41 (1-840) x AAY06948 (1-280)

QY 1 ATGAATTATAGAAAAATCTAGTAAGAAGCCGGTTAATCTCATTAAATGCTCAATCTTACCA 60
DB 1 MetAsnCysLysLysPhePheIleThrThrLeuValSerLeuMetSerPheLeuPro 20
QY 61 TATCAGTCTTTTGGAGATCCCTGTAGGTTCAAGAAGTAATGATAACAAAGAGGC---TTC 117
DB 21 GlyIleSerPheSerAspAlaVal-----GlnAsnAspAsnValGlyGlyAsnPhe 37
QY 118 TACATTAGTCAAAAGTACAATCCAAAGTATATACACATTTTAGAAAAATTTCTCTGCTGAAGA 177
DB 38 TyrIleSerGlyLysTyrValProSerValSerHisPheGlyValPheSerAlaLysGln 57
QY 178 ACTCCTATTAAATGAACAATAATCTCTCACTAAAAGTTTTCGAGCTAAAGAAAGATGCT 237
DB 58 -----GluArgAsnThrThrThrGlyValPheGlyLeuLysGlnAsnPro 72
QY 238 GAT-----ATAACAAAAAAGAC---CATTTTACAAGAGTAGCTCCAGGCAATTCAT 285
DB 73 AspGlySerThrIleSerLysAsnSerProGluAsnThrPheAsnValProAsnTyrSer 92
QY 286 TTT-----CAAAATAACTTAATATCATGAGATTTTTCAGGAAGTATTGGTTACTCTATGGAC 339
DB 93 PheLysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaValGlyTyrLeuMetAsn 112
QY 340 GGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGATAAC 399
DB 113 GlyProArgIleGluLeuGluMetSerTyrGluThrPheAspValLysAsnGlnGlyAsn 132
QY 400 AATGATCTGATAATGTTGAATACTATATAAACAATTTTGCATTATCTCGT-----447
DB 133 AsnTyrLysAsnAspAla-----HisLysTyrTyrAlaIleLeuThrHisAsnSerGlyGly 150
QY 448 AAAGATGCAATGGAAGATCAGCAATATGTAGTACTTAAAAATGACGGCATAACTTTTATG 507
DB 151 LysLeuSerAsnAlaGlyAspLysPheValPheLeuLysAsnGluGlyLeuLeuAspIle 170
QY 508 TCATTGATGTTAATACTTGTCTATGACATTCACAGCTGAGGAGTATCTTTTGTACCATAT 567
DB 171 SerLeuMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSerProTyr 190
QY 568 GCATGTGCAGGTATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAAAATTT 627
DB 191 IleCysAlaGlyValGlyThrAspLeuIleSerMetPheGluAlaIleAsnProLysIle 210
QY 628 GCTTACCAAGGAAAAATAGGTATTAGTTACCTATCACACCAAGAGTCTCTGCAATTTAT 687
DB 211 SerTyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSerValPheVal 230
QY 688 GGTGATACTACCATGGCTTATGCTTAATAAATTTGACAAGATACCTGTTAATAACTGCT 747
DB 231 GlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProAlaMetIlePro 250
QY 748 GTAGTATTAAATGATGCTCTCAACACCATCTGCTTCA-----GTAACCTCTT 795
DB 251 -----SerThrSerThrLeuThrGlyAsnHisPheThrIleValThrLeu 265
QY 796 GACGTGTGATACTTTGGCGGAGAAATTTGGAATGAGGTTTACCTTC 840
DB 266 SerValCysHisPheGlyValGluLeuGlyGlyArgPheAsnPhe 280

RESULT 20
AAU96110
ID AAU96110 standard; Protein; 280 AA.
XX
AC AAU96110;

```



PT animals  
 XX Claim 3; Fig 2B; 39pp; English.  
 PS This is the full-length variable surface antigen VSA4 protein of  
 CC Ehrlichia chaffeensis. Its amino acid sequence was deduced from a  
 CC partial open reading frame (ORF4) of a genomic locus (see AAV07179)  
 CC of E. chaffeensis that was obtained on the basis of homology to the  
 CC major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium.  
 CC This genomic locus included 5 ORFs encoding similar, but  
 CC non-identical proteins (see AAW51091-95). A claimed composition  
 CC comprises a nucleic acid (see AAV07176-82) encoding a polypeptide  
 CC (see AAW51088-99) that elicits a protective immune response against a  
 CC rickettsial pathogen. The nucleic acid is used, in human or  
 CC veterinary medicine, in vaccines to protect against Rickettsia,  
 CC Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic  
 CC polypeptides can be used diagnostically to detect antibodies  
 CC associated with Ehrlichia infection (claimed).  
 XX  
 SQ Sequence 280 AA;  
 Alignment Scores:  
 Pred. No.: 1,79e-59 Length: 280  
 Score: 603.00 Matches: 133  
 Percent Similarity: 63.3% Conservative: 54  
 Best Local Similarity: 45.0% Mismatches: 78  
 Query Match: 40.7% Indels: 30  
 DB: 19 Gaps: 10  
 US-10-062-624-41 (1-840) x AAW51094 (1-280)  
 QY 1 ATGAATTATACAAAATCTAGTAGAGCGGTTAAATCTCAATTAATGTCAATCTTACCA 60  
 DB 1 MetAsnCysLeuPhePheThrThrThrLeuValSerLeuMetSerPheLeuPro 20  
 QY 61 TATCAGTCTTTTGCAGATCTCTAGGTTCACGAACATAATGATACAAAGAGGC---TTC 117  
 DB 21 GlylleSerPheSerAspAlaVal-----GlnAsnAspAsnValGlyGlyAsnPhe 37  
 QY 118 TACATTAGTGCAGATACCAATCAAGTATATACACTTTAGAAAATTTCTGCTCAAGAA 177  
 DB 38 TyrlleSerGlyLysTyValProSerValSerHisPheGlyValPheSerAlaLysGln 57  
 QY 178 ACTCCTATATATGGAACAATTTCTCACTAAAAAGTTTCGGACTAAAGAAAGATGGT 237  
 DB 58 -----GluArgAsnThrThrIleGlyValPheGlyLeuLysGlnAspTrp 72  
 QY 238 GAT-----ATAACAAAAAGAC---GATTTTACAAGAGTAGCTCCAGGCATTGAT 285  
 DB 73 AspGlySerThrIleSerLysAsnSerProGluAsnThrPheAsnValProAsnTySer 92  
 QY 286 TTT-----CAAAATACTTAATATCAGGATTTTCAAGAGTATTTGGTTACTCTATGAC 339  
 DB 93 PheLysTyGluAsnAsnProPheLeuGlyPheAlaGlyAlaValGlyTyLeuMetAsn 112  
 QY 340 GGACCAAGATGACTTCACTGCATATCAACATTTAATCCAAAACACCCATAC 399  
 DB 113 GlyProArgIleGluLeuGluMetSerTyGluThrPheAspValLysAsnGlnGlyAsn 132  
 QY 400 AATGATCTGATAATGGTGAATCTATATAACATTTTGGCATTTATCTCGT----- 447  
 DB 133 AsnTyLysAsnAspAla-----HisLysTyTyValAlaLeuThrHisAsnSerGlyGly 150  
 QY 448 AAGATGCAATGGAAGATCAGCAATATGATGACTTAAATAATGACGGCATACTTTATG 507  
 DB 151 LysLeuSerAsnAlaGlyAspLysPheValPheLeuLysAsnGluGlyLeuLeuAspIle 170  
 QY 508 TCATTGATGTTAATCTGCTATGATACATTACAGCTGAGGAGTATCTTCATCATAT 567  
 DB 171 SerLeuMetLeuAsnAlaCysTySerAspValIleSerGluGlyIleProPheSerProTy 190  
 QY 568 GCATGTGAGGTATAGGACAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTT 627

Db 191 IleCysAlaGlyValGlyThrAspLeuIleSerMetPheGluAlaIleAsnProLysIle 210  
 QY 628 GCTTACCAAGAAAATAGTATTAGTATACCTATACACAGAGTCTCTGCAATTTATT 687  
 Db 211 SerTyGlnGlyLysLeuGlyLeuSerTySerIleSerProGluAlaSerValPheVal 230  
 QY 688 GGTGATACCTACCATGCGTTATTGGTAATAAATTTGAGAAGATACCTGTAAATACCTCT 747  
 Db 231 GlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProAlaMetIlePro 250  
 QY 748 GTAGTATTAAATGATGCTCTCTCAACACCATCTCTCTTCA-----GTAATCTTT 795  
 Db 251 -----SerThrSerThrLeuThrGlyAsnHisPheThrIleValThrLeu 265  
 QY 796 GACGTTGATATCTTTGGGAGAAATTTGGAATGAGGTTACCTTC 840  
 Db 266 SerValCysHisPheGlyValGluLeuGlyGlyArgPheAsnPhe 280  
 RESULT 22  
 AAB36188  
 ID AAB36188 standard; Protein; 280 AA.  
 XX AC AAB36188;  
 XX 02-MAR-2001 (first entry)  
 DT  
 XX Ehrlichia chaffeensis partial VSA4.  
 DE  
 KW Ehrlichia chaffeensis; VSA4; variable surface antigen 4; MAP1;  
 KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;  
 KW Rickettsia; Cowdria; Anaplasma; map2; lhwoif3; 4hwoif1; 18hwoif1;  
 KW 3gdorf3.  
 XX  
 OS Ehrlichia chaffeensis.  
 XX  
 PN WO200065063-A2.  
 XX  
 PD 02-NOV-2000.  
 XX  
 PF 21-APR-2000; 2000WO-US10886.  
 XX  
 PR 22-APR-1999; 99US-0130725.  
 XX  
 PA (UYFL ) UNIV FLORIDA.  
 XX  
 PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;  
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;  
 XX  
 DR WPI; 2000-679675/66.  
 XX N-PSDB; AAC68705.  
 XX  
 PT New polynucleotides useful as DNA vaccines for conferring immunity to  
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,  
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens  
 PT  
 PS Claim 3; Page 45-46; 63pp; English.  
 XX  
 CC The present sequence shows a high degree of similarity to the major  
 CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be  
 CC used in a vaccine to protect animals or humans against rickettsial  
 CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,  
 CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response  
 CC protective against the rickettsial pathogen. The nucleic acid vaccines  
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.  
 CC Cowdria ruminantium genes designated map 2, lhwoif3, 4hwoif1, 18hwoif1  
 CC and 3gdorf3 may be used in therapeutic and diagnostic applications. The  
 CC polypeptides are useful for detecting antibodies associated with  
 CC infection by a rickettsial pathogen whilst the polynucleotides may be  
 CC used to detect the presence of rickettsial nucleic acids.  
 XX  
 SQ Sequence 280 AA;

Alignment Scores:  
Pred. No.: 1.79e-59 Length: 280  
Score: 603.00 Matches: 133  
Percent Similarity: 63.39% Conservative: 54  
Best Local Similarity: 45.08% Mismatches: 78  
Query Match: 40.72% Indels: 30  
DB: 21 Gaps: 10

US-10-062-624-41 (1-840) x AAB36188 (1-280)

Qy 1 ATGAATTATAAGAAATTCCTAGAGCGGTTAATCTCATTAAATGCTCAATCTTACCA 60  
Db 1 MetAsnCysLysLysPhePheIleThrThrLeuValSerLeuMetSerPheLeuPro 20  
Qy 61 TATCAGCTTTTGCAGATCTCTGAGTTCAAGAACTAATGATAACAAGAGGC---TTC 117  
Db 21 GlyIleSerPheSerAspAlaVal-----GlnAsnAspAsnValGlyGlyAsnPhe 37  
Qy 118 TACATTAGTCAAGTACCAATCCAAGTATATCACACTTTAGAAAATCTCTGCTGAAGAA 177  
Db 38 TyrIleSerGlyLysTyTyValProSerValSerHisPheGlyValPheSerAlaLysGln 57  
Qy 178 ACTCCTATTATGAACAAATCTCTCACTAAAGTTTCGGACTAAAGAAAGATGGT 237  
Db 58 -----GluArgAsnThrThrIleGlyValPheGlyLeuLysGlnAspTrp 72  
Qy 238 GAT-----ATAACAAAAAAGAC---GATTTTACAAGAGTAGCTCCAGGCGATTGAT 285  
Db 73 AspGlySerThrIleSerLysAsnSerProGluAsnThrPheAsnValProAsnTySer 92  
Qy 286 TTT-----CAAAATACTTAATATCAGGATTTTCAGGAAGTATGTTACTCTATGAC 339  
Db 93 PheLysTyTyGluAsnAsnProPheLeuGlyPheAlaGlyAlaValGlyTyTyLeuMetAsn 112  
Qy 340 GGACCAAGATAGAACTTGAAGTGCATATCAACAATTTAATCCAAAAACACCGATAAC 399  
Db 113 GlyProArgIleGluLeuGluMetSerTyTyGluThrPheAspValLysAsnGlnGlyAsn 132  
Qy 400 AATGATAGTATAATGTTGAATACTATAAATTTGCTATATCTCGT-----447  
Db 133 AsnTyTyLysAsnAspAla-----HisLysTyTyTyAlaLeuThrHisAsnSerGlyGly 150  
Qy 448 AAGATGCAATGGAAGATCAGCAATATGTAGTACTTAAATATGAGCGCATACCTTTATG 507  
Db 151 LysLeuSerAsnAlaGlyAspLysPheValPheLeuLysAsnGluGlyLeuLeuAspIle 170  
Qy 508 TCATTGATGGTTAATCTACTGTGTATGACATTCAGCTGAGGAGTATCTTCGTACCATAT 567  
Db 171 SerLeuMetLeuAsnAlaCysTyAspValIleSerGluGlyIleProPheSerProTy 190  
Qy 568 GCATGTGAGTATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAAATTT 627  
Db 191 IleCysAlaGlyValGlyThrAspLeuIleSerMetPheGluAlaIleAsnProLysIle 210  
Qy 628 GCTTACCAAGAAAAATAGTATTAGTTACCTATCACACAGAGTCTCTGCAATTTAT 687  
Db 211 SerTyTyGlnGlyLysLeuGlyLysSerTyTySerIleSerProGluAlaSerValPheVal 230  
Qy 688 GGTGATACCTACCATGGCTTATGTTATTAATAATTTGGAAGATACCTGTAATACCTCT 747  
Db 231 GlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProAlaMetIlePro 250  
Qy 748 GTAGTATTAAATGATGCTCTCAACACACATCTCTCTCA-----GTAACCTCTT 795  
Db 251 -----SerThrSerThrLeuThrGlyAsnHisPheThrIleValThrLeu 265  
Qy 796 GACCTTGATATCTTGGCGGAGAAATTTGGAATGAGTTACCTTC 840  
Db 266 SerValCysHisPheGlyValGluLeuGlyGlyArgPheAsnPhe 280

RESULT 23  
AAU04198  
ID AAU04198 standard; Protein; 280 AA.

XX AAU04198;  
AC 23-OCT-2001 (first entry)  
DT Variable surface antigen 4 (VSA4) from *Ehrlichia chaffeensis*.  
DE Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;  
KW infection; heartwater; diagnostic; variable surface antigen; VSA.  
XX *Ehrlichia chaffeensis*.  
OS US6251872-B1.  
FN 26-JUN-2001.  
PD 17-OCT-1997; 97US-0953326.  
PF 17-OCT-1996; 96US-0733230.  
PR (UYFL) UNIV FLORIDA.  
XX Barbet AF, Ganta RR, McGuire TC, Burrige MJ, Nyika A;  
PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;  
XX WPI; 2001-424487/45.  
DR N-PSDB; AAS07578.  
XX New MAP2 genes and polypeptides useful as vaccines for conferring  
PT immunity to human and animal rickettsial diseases, e.g. heartwater, or  
XX as molecular markers in nucleic acid analysis procedures -  
PS Example 3; Fig 2A-2B; 30pp; English.  
XX The sequence represents the amino acid sequence of variable surface  
CC antigen 4 (VSA4) isolated from *Ehrlichia chaffeensis*, which  
CC has similarity to major antigen protein (MAP). The MAP polynucleotides  
CC and polypeptides are useful as vaccines for conferring immunity to  
CC rickettsia infection, including *Cowdria ruminantium* causing heartwater.  
CC The MAP polynucleotides may be used as molecular markers in nucleic acid  
CC analysis procedures, and to produce the MAP polypeptides, which may  
CC be used to raise antibodies that are reactive with the polypeptides.  
CC The nucleic acids may further be used as probes to identify  
CC complementary sequences within other nucleic acid molecules or genomes,  
CC where such probes can be applied to identify or distinguish infectious  
CC strains of organisms in diagnostic procedures or in rickettsial  
CC research where identification of particular organisms or strains is  
CC needed.  
SQ Sequence 280 AA;  
Alignment Scores:  
Pred. No.: 1.79e-59 Length: 280  
Score: 603.00 Matches: 133  
Percent Similarity: 63.39% Conservative: 54  
Best Local Similarity: 45.08% Mismatches: 78  
Query Match: 40.72% Indels: 30  
DB: 21 Gaps: 10  
US-10-062-624-41 (1-840) x AAU04198 (1-280)  
Qy 1 ATGAATTATAAGAAATTCCTAGAGCGGTTAATCTCATTAAATGCTCAATCTTACCA 60  
Db 1 MetAsnCysLysLysPhePheIleThrThrLeuValSerLeuMetSerPheLeuPro 20  
Qy 61 TATCAGCTTTTGCAGATCTCTGAGTTCAAGAACTAATGATAACAAGAGGC---TTC 117  
Db 21 GlyIleSerPheSerAspAlaVal-----GlnAsnAspAsnValGlyGlyAsnPhe 37  
Qy 118 TACATTAGTCAAGTACCAATCCAAGTATATCACACTTTAGAAAATCTCTGCTGAAGAA 177  
Db 38 TyrIleSerGlyLysTyTyValProSerValSerHisPheGlyValPheSerAlaLysGln 57





```

FT Peptide 1..23
FT /label= Signal_peptide
FT Protein 24..280
XX /label= Mature_ECa28SA3_28-kDa_protein
XX WO200032745-A2.
XX
XX 08-JUN-2000.
XX
XX 24-NOV-1999; 99WO-US28075.
XX
XX 30-NOV-1998; 98US-0201458.
XX 03-MAR-1999; 99US-0261358.
XX
XX (RERE-) RES DEV FOUND.
XX
XX Walker DH, Yu X, McBride JW;
XX WPI; 2000-412298/35.
XX DR N-PSDB; AAD01294, AAD01295.
XX
XX Ehrlichia canis antigens useful for vaccinating against canine
XX ehrlichiosis in dogs -
XX
XX Claim 12; Page 68-69; 86pp; English.
XX
XX The patent relates to homologous 28-kilodalton (kDa) protein genes of
XX Ehrlichia canis, designated Eca28SA1, Eca28SA2, Eca28-1 and
XX Eca28-2. These genes are members of a polymorphic multiple gene family
XX and contained in a single locus of 5.592 kb. The 28-kDa proteins are
XX immunoreactive with anti-E. canis serum hence are important
XX immunoprotective antigens. The protein is useful for vaccinating
XX against E. canis infections such as canine ehrlichiosis in dogs.
XX Canine ehrlichiosis, also known as canine tropical pancytopenia, is a
XX tick-borne rickettsial disease of dogs. Eca28-1 is conserved amongst
XX different strains of E. canis and hence useful for serodiagnosis of
XX canine ehrlichiosis. The present sequence is a E. canis
XX Eca28SA3 30-kDa protein which is post-translationally modified to a
XX mature 28-kDa protein by cleavage of N-terminal signal sequence.
XX
SQ Sequence 280 AA;

Alignment Scores:
Pred. No.: 3 01e-59 Length: 280
Score: 601.00 Matches: 129
Percent Similarity: 61.59% Conservative: 49
Best Local Similarity: 44.64% Mismatches: 93
Query Match: 40.58% Indels: 18
DB: 21 Gaps: 6

US-10-062-624-41 (1-840) x AAY71479 (1-280)

QY 1 ATGAATTATAGAAATCTCTAGTAAGACGGCTTAATCTCTAATTAATGTCAATCTTACCA 60
DB 1 MetAsnCysLeuLysLeuLeuLeuThrAlaLeuMetSerLeuMetTyrTyrAlaPro 20
QY 61 TATAGTCTTTTGCAGATCTCTAGGTTCAAGAACTAATGATTAACAAGAGCTTCTAC 120
DB 21 SerIleSerPheSerAspThrIle-----GlnAspAspAsnThrGlySerPheTyr 37
QY 121 ATTAGTGCAGAACTACAATCAAGTATATCACACTTTAGAAAAATTTCTCTGCTGAAGAACT 180
DB 38 IleSerGlyLysTyrValProSerValSerHisPheGlyValPheSerAlaLysGlu--- 56
QY 181 CCTATTAAATGGAAACAAATCTCTCACTAAATAAGTTTTCGGACTAAAGAAAGAT----- 234
DB 57 -----GluArgAsnSerThrValGlyValPheGlyLeuLysHisAspTrpAsn 72
QY 235 ---GCTGATATACAAAA-----AAAGACGATTTTACAAGAGTAGCTCCAGGCATT 282
DB 73 GlyGlyThrIleSerAsnSerSerProGluAsnIlePheThrValGlnAsnTyrSerPhe 92
QY 283 GATTTCAAAATAACTTAATATATCAGGATTTTCAGGAAGTATTGGTTACTTCTATCGACGGA 342

```

---

```

Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGlyGly 112
QY 343 CCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACACGATAACAAT 402
Db 113 ProArgIleGluLeuGluValLeuTyrPheAspValLysAsnGlnAsnAsn 132
QY 403 GATACTGATAATGCTGAATACATAT-----AAACATTTTGCATTATCTCGTAAAGAT 453
Db 133 TyrLysAsnGlyAlaHisArgTyrCysAlaLeuSerHisHisSerSerAlaThrSerMet 152
QY 454 GCAATGGAAGATCAGCAATATGTACTTAAAAATGACGGCATAACTTTTATGTCAATTG 513
Db 153 SerSerAlaSerAsnLysPheValPheLeuLysAsnGluGlyLeuIleAspLeuSerPhe 172
QY 514 ATGGTTAATACTTGTCTATGACATTACAGCTCAAGAGATATCTTTCGTACCATATGTCATGT 573
Db 173 MetIleAsnAlaCysTyrAspIleIleIleGluGlyMetProPheSerProTyrIleCys 192
QY 574 GCAGGTATAGGACGAGATCTTATCATCTATTTTAAAGACCTCAATCTAAAAATTTGCTTAC 633
Db 193 AlaGlyValGlyThrAspValValSerMetPheGluAlaIleAsnProLysIleSerTyr 212
QY 634 CAAGGAAAAATAGGTATTAGTTACCTATCACACAGAAAGTCTCTGCAATTTATTGGTGA 693
Db 213 GlnGlyLysLeuGlyLeuGlyTyrSerIleSerSerGluAlaSerValPheIleGlyGly 232
QY 694 TACTACATGCGGTTTATGGTAAATTTGAGAGATACCTGTAACTCTCTGTAGTA 753
Db 233 HisPheHisArgValIleGlyAsnGluPheArgAspIleProAlaMetValProSerGly 252
QY 754 TTAATGATGCTCTCTCAACACACATCTGCTTTCAGTAACTCTTGAGTTGGATACCTTGC 813
Db 253 SerAsn---LeuProGluAsnGlnPheAlaIleValThrLeuAsnValCysHisPheGly 271
QY 814 GGAGAAATTTGAATGAGGTTCACCTTC 840
Db 272 IleGluLeuGlyGlyArgPheAsnPhe 280

RESULT 27
AAU96102
ID AAU96102 standard; Protein; 280 AA.
XX
XX AC AAU96102;
XX
XX DT 02-JUL-2002 (first entry)
XX
XX DE Ehrlichia canis p28-6.
XX
XX KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;
XX antibacterial.
XX
XX OS Ehrlichia canis.
XX
XX PN WO200222782-A2.
XX
XX PD 21-MAR-2002.
XX
XX PF 12-SEP-2001; 2001WO-US28759.
XX
XX PR 12-SEP-2000; 2000US-0660587.
XX
XX PA (RERE-) RES DEV FOUND.
XX
XX PI Walker DH, Yu X, McBride JW;
XX
XX DR WPI; 2002-351882/38.
XX
XX DR N-PSDB; ASK68854.
XX
XX New recombinant homologous 28 kilodalton immunodominant protein from
XX Ehrlichia canis, useful for treating Ehrlichia canis infections -
XX
XX Claim 16; Figure 7; 106pp; English.

```







```

Db 212 VallysPheAlaHisGlnGlyLysValGlyIleSerTySerIleSerProGluValSer 231
QY GCATTTATTGGTGATACCTACCATCGCGTTATTGGTAATAATTTGAGAAGATACCTGTA 738
Db 232 LeuPheLeuAenGlyTyTyRHisLysValThrGlyAenAArgPheLysAenLeuHisVal 251
QY 739 ATAACCTCTAGTATTAAATGATCTCTCAACACCAATCTGCTTCAGTAACTCTTGAC 798
Db 252 GlnHisValSerAspLeuSerAspAlaProLysPheThrSerAlaValAlaThrLeuAen 271
QY 799 GTTGCACTACTTGGCGGAGAAATTCGAATGAGGTTACCTTC 840
Db 272 ValGlyTyRPhGlyGlyGluIleGlyValArgPheIlePhe 285

```

## RESULT 30

AA06961  
ID AAY06961 standard; Protein; 307 AA.

AC AAY06961;

DT 05-JUL-1999 (first entry)

DE E. canis P30-1 protein.

XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
KW detection; dog.

XX Ehrlichia canis.

OS WO9913720-A1.

FN 25-MAR-1999.

XX 18-SEP-1998; 98WO-US19600.

PF 19-SEP-1997; 97US-0059353.

PR (OHIS ) UNIV OHIO STATE.

PA Ohashi N, Rikihisa Y;

PI WPI, 1999-254290/21.

XX N-PSDB; AAX34761.

XX Novel outer membrane proteins from Ehrlichia chaffeensis and  
PT Ehrlichia canis

PS Disclosure; Fig 21B; 55pp; English.

XX The invention provides isolated outer membrane proteins (OMP) from  
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
CC in AAY06943-958. The E. canis proteins form part of the P30 family and  
CC consist of proteins shown in AAY06959-970. The proteins and genes are  
CC used to detect E. chaffeensis in patients and E. canis in dogs.

XX Sequence 307 AA;

## Alignment Scores:

Pred. No.:	9.28e-58	Length:	307
Score:	588.00	Matches:	131
Percent Similarity:	59.93%	Conservative:	41
Best Local Similarity:	45.64%	Mismatches:	99
Query Match:	39.70%	Indels:	16
DB:	20	Gaps:	4

US-10-062-624-41 (1-840) x AAY06961 (1-307)

QY 1 ATGAATTATAGAAATTTCTAGTAGAAGCGGTTAATCTCATTAATGCTCAATCTTACCA 60

Db 30 MetaAenCysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 49

```

QY 61 TATCAGTCTTTTGGCAGATCCTGTAGGTTCAAGAACTAATGTATATAACAAGAAGGC----- 114
Db 50 SerIleSerPheSerAspThrIle-----GlnAepGlyAenMetGlyGlyAen 65
QY 115 TTCTACATTAGTGCAGAACTCAATCAAGTATATACACATTTAGAAAATTTCTCTGCTGAA 174
Db 66 PheTyRleSerGlyLysTyRValProSerValSerHisPheGlySerPheSerAlaLys 85
QY 175 GAAACTCTCTATTAAATGAAACAAATTTCTCTCACTAAAAAGTTTTCGGAGCTAAAGAAAGAT 234
Db 86 Glu-----GluSerLysSerThrValGlyValPheGlyLeuLysHisAsp 100
QY 235 GGTGAT-----ATAACAAAAAGACGATTTTACAAGTAGTCCAGGC 279
Db 101 TrpAspGlySerProIleLeuLysAenLysHisAlaAspPheThrValProAenTyRser 120
QY 280 ATTGATTTTCAAAATAACTTAATATCAGGATTTTTCAGGAAGTATTTGTTACTCTATGGAC 339
Db 121 PheArgTyRGlueAenAenProPheLeuGlyPheAlaGlyAlaIleGlyTyRserMetGly 140
QY 340 GGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTTATCCAAAAAACACCGATAC 399
Db 141 GlyProArgIleGluPheGluIleSerTyRGlueAlaPheAspValLysSerProAenIle 160
QY 400 AATGATCTGATAATGTTGTAATCTATAAATTTTTCGATTTATCTCGTAAAGATGCAATG 459
Db 161 AenTyRGlueAenAspAlaHisArgTyR-CysAlaLeuSerHisHisThrSerAlaAlaMet 180
QY 460 GAAGTACGCAATATAGTACTTAAATAATCAGCGCATACCTTTTATGCTATTGATGTT 519
Db 181 GluAlaAspLysPheValPheLeuLysAenGluGlyLeuIleAspIleSerLeuAlaIle 200
QY 520 AATACTTGTCTATGACATTACAGCTGAAGAGGATCTTTCGTACCATCATGTCAGTGCAGG 579
Db 201 AsnAlaCysTyRAspIleIleAsnAspLysValProValSerProTyRlleCysAlaGly 220
QY 580 ATAGGACAGATCTTATCATCTATTTTTAAAGACCTCAATCTTAAATTTGCTTACCAAGA 639
Db 221 IleGlyThrAspLeuIleSerMetPheGluAlaThrSerProLysIleSerTyRGlueGly 240
QY 640 AAAATAGTATTAGTTACCTTACCATCAGCGAGTCTCTGCATTTATTTGGTGGATAC 699
Db 241 LysLeuGlyIleSerTyRserIleAsnProGluThrSerValPheIleGlyHisPhe 260
QY 700 CATGCGCTTATTGGTAATAATTTGAGAAGATACCTGTAATACTCTGTAGTATTAAAT 759
Db 261 HisArgIleIleGlyAenGluPheArgAspIleProAlaIleValProSerAsnSerThr 280
QY 760 GATGCTCTCAACCAACATCTGCTTCAAGTAACTCTTGCAGCTTGGATATCTTTGGCGGAGAA 819
Db 281 ThrIleSerGlyProGlnPheAlaThrValThrLeuAenValCysHisPheGlyLeuGlu 300
QY 820 ATTGGAATGAGTTTCACCTTC 840
Db 301 LeuGlyGlyArgPheAenPhe 307
RESULT 31
AAY71477
ID AAY71477 standard; Protein; 278 AA.
XX AC AAY71477;
XX DT 12-OCT-2000 (first entry)
XX DE Ehrlichia canis immunoreactive protein ECA28-1.
XX KW Homologous 28-kDa protein gene; ECA28-1; immunoreactive; vaccine;
XX KW p28 gene; polymorphic multiple gene family; immunoprotective antigen;
XX KW antibacterial; canine ehrlichiosis; canine tropical pancytopenia;
XX KW tick-borne rickettsial disease; serodiagnosis.
XX OS Ehrlichia canis.
XX

```

```

FH Key Location/Qualifiers
FT Peptide 1..23
FT Protein /label= Signal_peptide
FT Cleavage-site 21..23
FT Cleavage-site /label= Mature_ECa28-1_28-kDa_protein
FT Cleavage-site 23..25
FT Cleavage-site /label= Signal_peptidase_cleavage_site
FT Region /note= "Additional putative signal peptidase cleavage site"
FT Region 27..34
FT Region /label= Variable_region_1
FT Region /label= "contains immunoreactive peptides"
FT Region 76..84
FT Region /label= Variable_region_2
FT Region /note= "contains immunoreactive peptides"
FT Region 148..156
FT Region /label= Variable_region_3
FT Region /note= "contains immunoreactive peptides"
FT Region 246..258
FT Region /label= Variable_region_4
FT Region /note= "contains immunoreactive peptides"
XX WO200032745-A2.
XX 08-JUN-2000.
XX 24-NOV-1999; 99WO-US28075.
XX 30-NOV-1998; 98US-0201458.
XX 03-MAR-1999; 99US-0261358.
XX (RERE-) RES DEV FOUND.
XX Walker DH, Yu X, McBride JW;
XX WPI: 2000-412298/35.
XX N-PSDB; AAD01292.
XX Ehrlichia canis antigens useful for vaccinating against canine ehrlichiosis in dogs -
XX Claim 12; Fig 1; 86pp; English.
XX The patent relates to homologous 28-kilodalton (kDa) protein genes of Ehrlichia canis, designated Eca28A1, Eca28A2, Eca28A3, Eca28-1 and Eca28-2. These genes are members of a polymorphic multiple gene family and contained in a single locus of 5.92 kb. The 28-kDa proteins are immunoreactive with anti-E. canis serum hence are important immunoprotective antigens. The protein is useful for vaccinating against E. canis infections such as canine ehrlichiosis in dogs. Canine ehrlichiosis, also known as canine tropical pancytopenia, is a tick-borne rickettsial disease of dogs. Eca28-1 is conserved amongst different strains of E. canis and hence useful for serodiagnosis of canine ehrlichiosis. The present sequence is a E. canis Eca28-1 30-kDa protein which is post-translationally modified to a mature 28-kDa protein by cleavage of N-terminal signal sequence.
XX Sequence 278 AA;

Alignment Scores:
Pred. No.: 4,3e-57 Length: 278
Score: 582.00 Matches: 131
Percent Similarity: 59.93% Conservative: 41
Best Local Similarity: 45.64% Mismatches: 99
Query Match: 39.30% Indels: 16
DB: 21 Gaps: 4

US-10-062-624-41 (1-840) x AAY71477 (1-278)

QY 1 ATGAATTATAGAAATCTAGTAAGACGGCGTTAATCTATTATGTCAATCTTACCA 60
Db 1 MetAsnCysLeuLysLeuLeuLeuThrAlaLeuLeuSerLeuMetTyrSerIlePro 20

```

```

QY 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAAGAACTAATGATATAACAAGAAGGC----- 114
Db 21 SerIleSerPheSerAspThrIle-:-----GlnAspGlyAsnMetGlyGlyAsn 36
115 TTCTACATTAGTGCAAGTACAATCCAAAGTATATACACATTTAGAAAAATCTCTGCTGAA 174
Db 37 PheTyrIleSerGlyLysTyrValProSerValSerHisPheGlySerPheSerAlaLys 56
175 GAAACTCTCTATTAAATGAACAATAATCTCTCACTAAAAAGTTTTCGGACTTAAGAAAGAT 234
Db 57 Glu-:-----GluSerLysSerThrValGlyValPheGlyLeuLysHisAsp 71
235 GGTGAT-----ATAACAATAAAGACGATTTTACAAGAGTAGTCCAGGC 279
Db 72 TrpAspGlySerProIleLeuLysAsnLysHisAlaAspPheThrValProAsnTyrSer 91
280 ATTGATTTTCAAAATAACTTAATATCAGGATTTTTCAGAGAGTATTTGTTACTTCTATGAC 339
Db 92 PheArgTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGly 111
340 GGACCAAGATPAGAACTTGAACTGCATATCAACAATTTTAAATCCAAAAACACCGATAAC 399
Db 112 GlyProArgIleGluPheGluIleSerTyrGluAlaPheAspValLysSerProAsnIle 131
400 AATGATCTGATAATGTGTAATACTATATAACATTTTGCATTTATCTCGTAAAGATGCAATG 459
Db 132 AsnTyrGlnAsnAspAlaHisArgTyrCysAlaLeuSerHisHisThrSerAlaAlaMet 151
460 GAAGATCAGCAATATGTAGTACTTTAAAATGACGGCATACTTTTATGTCTATGATGTT 519
Db 152 GluAlaAspLysPheValPheLeuLysAsnGluGlyLeuIleAspIleSerLeuAlaIle 171
520 AATACTTGTCTATGACATTTACAGCTGAAGAGTATCTTTTCGTACCATATGTCAGGT 579
Db 172 AsnAlaCysTyrAspIleIleAsnAspLysValProValSerProTyrIleCysAlaGly 191
580 ATAGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAGGA 639
Db 192 IleGlyThrAspLeuIleSerMetPheGluAlaThrSerProLysIleSerTyrGlnGly 211
640 AAAATAGTATTAGTTACCTTATCACACCAAGCTCTCTGCATTTATTTGGTGGATACCTAC 699
Db 212 LysLeuGlyIleSerTyrSerIleAsnProGluThrSerValPheIleGlyGlyHisPhe 231
700 CATCGCGTTATTGTAATAAAATTTGAGAAGTACTGTATAACTCTCTGTAGTATTAAAT 759
Db 232 HisArgIleIleGlyAsnGluPheArgAspIleProAlaIleValProSerAsnSerThr 251
760 GATGCTCTCAACACCATCTGCTTACAGTCTTGCAGTCTTGCAGTCTTGGATCTTTGGCGGAA 819
Db 252 ThrIleSerGlyProGlnPheAlaThrValThrLeuAsnValCysHisPheGlyLeuGlu 271
820 ATTGGAATGAGGTTTCACCTTC 840
Db 272 LeuGlyGlyArgPheAsnPhe 278

RESULT 32
AAU96100
ID AAU96100 standard; Protein; 278 AA.
XX AC AAU96100;
XX DT 02-JUL-2002 (first entry)
XX DE Ehrlichia canis p28-7.
XX KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;
XX OS Ehrlichia canis.
XX PN WO200222782-A2.

```



CC non-identical proteins (see AAWS1091-95). A claimed composition  
 CC comprises a nucleic acid (see AAV07176-82) encoding a polypeptide  
 CC (see AAWS1088-99) that elicits a protective immune response against a  
 CC rickettsial pathogen. The nucleic acid is used, in human or  
 CC veterinary medicine, in vaccines to protect against Rickettsia,  
 CC Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic  
 CC polypeptides can be used diagnostically to detect antibodies  
 CC associated with Ehrlichia infection (claimed).

XX SQ Sequence 278 AA;

Alignment Scores:  
 Pred. No.: 1,596-56 Length: 278  
 Score: 577.00 Matches: 126  
 Percent Similarity: 60.55% Conservative: 49  
 Best Local Similarity: 43.60% Mismatches: 94  
 Query Match: 38.96% Indels: 20  
 DB: 19 Gaps: 7

US-10-062-624-41 (1-840) x AAWS1093 (1-278)

QY 1 ATGAATTATAAGAAATTTCTAGTAAGACGGGTTAATCTCTTAATCTCAATCTTACCA 60  
 Db 1 MetAenCysLysPhePheleThrAlaLeuValSerLeuMetSerPheLeuPro 20  
 QY 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAGAACTAATGATAACAAGAGGC---TTC 117  
 Db 21 GlylleSerPheSerAspProVal-----GlnGlyAspAsnIleSerGlyAsnPhe 37  
 QY 118 TACATTAGTCAAGTACAACTCAAGTATATCACACTTTAGAAAATTTCTCTCGAAGAA 177  
 Db 38 TyrValSerGlyLysTyrMetProSerAlaSerHisPheGlyMetPheSerAlaLysGlu 57  
 QY 178 ACTCTATTAAATGGAACAATTTCTCTACTAAAGATTTTCGGACTAAAGAAAGAT--- 234  
 Db 58 -----GluLysAsnProThrValAlaLeuTyrGlyLeuLysGlnAspTrp 72  
 QY 235 -----GGTGATATACAAAAGAGAGGATTTTACAGAGTAGTCCAGGCATT 282  
 Db 73 GluGlylleSerSerHisAsnAspAsnHisPheAsnAsnLysGlyTyrSerPhe 92  
 QY 283 GATTTTCAAAATAACTTAATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGACGGA 342  
 Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGlyGly 112  
 QY 343 CCAAGATAAGACTTGAAGCTGCATATCAACAATTTAATTCAAAAAACCCGATAACAA 402  
 Db 113 ProArgValGluPheGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn 132  
 QY 403 GATCTGATAAGGTGAATACATATATAACATTTTGCATTTATCTCGTAAGAT----- 453  
 Db 133 TyrLysAsnAspAla-----HisArgTyrCysAlaLeuGlyGlnGlnAspAsnSerGly 150  
 QY 454 GCAATGGAAGATCAGCAATATGTAGTACTTAAATAATGACGGCATACTTTTATGTATG 513  
 Db 151 IleProLysThrSerLysTyrValLeuLeuLysSerGluGlyLeuLeuAspIleSerPhe 170  
 QY 514 ATGGTTAATPACTTGTATGACATTTACAGCTGGAAGGAGTATCTTCGTACCATATGCAT 573  
 Db 171 MetLeuAsnAlaCysTyrAspIlelleAsnGluSerIleProLeuSerProTyrIleCys 190  
 QY 574 GCAGGTATAGGACAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTAC 633  
 Db 191 AlaGlyValGlyThrAspLeulleSerMetPheGluAlaThrAsnProLysIleSerTyr 210  
 QY 634 CAAGGAAAATAGGTATTAGTTACCCCTATCACACAGAGTCTCTCATTTATTGGTGA 693  
 Db 211 GlnGlyLysLeuGlyLeuSerTyrSerIleAsnProGluAlaSerValPheIleGlyGly 230  
 QY 694 TACTACCATGGCTTATTGGTATAATAATTTGAGAAGATACCTGTGTAATAACTCCTGTG 753  
 Db 231 HisPheHisLysValIleGlyAsnGluPheArgAspIleProThrLeuLysAlaPheVal 250

QY 754 TTAATGATGCTCTCTCAACACACATCTGCTTCACTAATCTCTTACGTTGGATCTTGGC 813  
 Db 251 ThrSerSerAlaThrProAspLeu--AlaIleValThrLeuSerValCysHisPheGly 269  
 QY 814 GGAGAAATTTGAATGAGGTTTCACCTTC 840  
 Db 270 IleGluLeuGlyGlyArgPheAsnPhe 278

RESULT 34

AAV06947  
 ID AAY06947 standard; Protein; 278 AA.

XX AC AAY06947;  
 XX 05-JUL-1999 (first entry)  
 XX E. chaffeensis OMP-1E protein.  
 XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 KW detection; dog.

XX OS Ehrlichia chaffeensis.

XX PN WO9913720-A1.

XX PD 25-MAR-1999.

XX PF 18-SEP-1998; 98WO-US19600.

XX PR 19-SEP-1997; 97US-0059353.

XX PA (OHIS ) UNIV OHIO STATE.

XX PI Ohashi N, Rikihisa Y;

XX DR WPI; 1999-254290/21.  
 N-PSDB; AAX34747.

XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
 Ehrlichia canis

XX PS Claim 15; Fig 7B; 55pp; English.

CC The invention provides isolated outer membrane proteins (OMP) from  
 Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 in AAY06943-958. The E. canis proteins form part of the P30 family and  
 consist of proteins shown in AAY06959-970. The proteins and genes are  
 used to detect E. chaffeensis in patients and E. canis in dogs.

XX SQ Sequence 278 AA;

Alignment Scores:  
 Pred. No.: 1,596-56 Length: 278  
 Score: 577.00 Matches: 126  
 Percent Similarity: 60.55% Conservative: 49  
 Best Local Similarity: 43.60% Mismatches: 94  
 Query Match: 38.96% Indels: 20  
 DB: 19 Gaps: 7

US-10-062-624-41 (1-840) x AAY06947 (1-278)

QY 1 ATGAATTATAAGAAATTTCTAGTAAGACGGGTTAATCTCTTAATCTCAATCTTACCA 60  
 Db 1 MetAenCysLysLysPhePheleThrAlaLeuValSerLeuMetSerPheLeuPro 20  
 QY 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAGAACTAATGATAACAAGAGGC---TTC 117  
 Db 21 GlylleSerPheSerAspProVal-----GlnGlyAspAsnIleSerGlyAsnPhe 37  
 QY 118 TACATTAGTCAAGTACAACTCAAGTATATCACACTTTAGAAAATTTCTCTCGAAGAA 177  
 Db 38 TyrValSerGlyLysTyrMetProSerAlaSerHisPheGlyMetPheSerAlaLysGlu 57







Db 251 ThrSerSerAlaThrProAspLeu---AlaIleValThrLeuSerValCysHiePheGly 269  
 Qy 814 GGAGAAATTGGAATGAGGTTCACTTC 840  
 Db 270 IleGluLeuGlyGlyArgPheAsnPhe 278

RESULT 37  
 AAU96109  
 ID AAU96109 standard; Protein; 278 AA.  
 AC AAU96109;  
 XX  
 XX 02-JUL-2002 (first entry)  
 DT  
 XX Ehrlichia chaffeensis OMP-1E.  
 DE  
 DE Ehrlichia canis infection; vaccine; serodiagnostic; p28;  
 KW  
 KW antibacterial.  
 KW  
 XX Ehrlichia chaffeensis.  
 OS  
 XX  
 XX WO200222782-A2.  
 PN  
 XX  
 XX 21-MAR-2002.  
 PD  
 XX  
 XX 12-SEP-2001; 2001WO-US28759.  
 PF  
 XX  
 XX 12-SEP-2000; 2000US-0660587.  
 PR  
 XX  
 XX (RERE-) RES DEV FOUND.  
 PA  
 XX  
 XX Walker DH, Yu X, McBride JW;  
 PI  
 XX  
 XX WPI; 2002-351882/38.  
 DR  
 XX  
 XX New recombinant homologous 28 kilodalton immunodominant protein from  
 PT Ehrlichia canis, useful for treating Ehrlichia canis infections -  
 PT  
 XX  
 XX Example 3; Figure 3; 106pp; English.  
 PS  
 XX  
 XX The invention relates to a recombinant homologous 28 kDa immunodominant  
 CC protein, p28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably  
 CC dispersed in a pharmaceutically acceptable carrier, is useful for  
 CC inhibiting E. canis infection in a subject. (I) is useful in the  
 CC development of vaccines and serodiagnostics that are particularly  
 CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118  
 CC represent the 28-kDa antigen amino acid sequences of the invention.  
 CC  
 XX  
 XX SQ Sequence 278 AA;

Alignment Scores:  
 Pred. No.: 1.59e-56 Length: 278  
 Score: 577.00 Matches: 126  
 Percent Similarity: 60.55% Conservative: 49  
 Best Local Similarity: 43.60% Mismatches: 94  
 Query Match: 38.96% Indels: 20  
 DB: 23 Gaps: 7

US-10-062-624-41 (1-840) x AAU96109 (1-278)

Qy 1 ATGAATTATAAGAAATTCCTAGTAAGAGCGGTTAATCTCAATTAATGCTCAATCTTACCA 60  
 Db 1 MetAsnCysLysLysPhePheIleThrAlaLeuValSerLeuMetSerPheLeuPro 20  
 Qy 61 TATCAGTCTTTTGCAGATCCCTGAGTTCAGAGACTAATGATAACAAGAAGGC---TTC 117  
 Db 21 GlytleSerPheSerAspProVal-----GlnGlyAspAsnIleSerGlyAsnPhe 37  
 Qy 118 TACATTAGTCAAAAGTACAATCCAAAGTATATACACACTTTTAGAAAAATCTCTGCTGGAAGAA 177  
 Db 38 TyrValSerGlyLysTyrMetProSerAlaSerHisPheGlyMetPheSerAlaLysGlu 57  
 Qy 178 ACTCCTATTAAATGAACAAATTCCTCTCACTAAAAAAGTTTTCGGACTAAAGAAGAT--- 234

Db 58 -----GluLysAsnProThrValAlaLeuTyrGlyLeuLysGlnAspTrp 72  
 Qy 235 -----GGTGATATAACAAAAAGACGATTTTACAAGAGTAGTCCAGGCATT 282  
 Db 73 GluGlyIleSerSerSerHisAsnAspAsnHisPheAsnAsnLysGlyTyrSerPhe 92  
 Qy 283 GATTTCAAAATAAATTAATATCAGGATTTTCAGGAAGTATTGGTTACTCTTATGGACGGA 342  
 Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGlyGly 112  
 Qy 343 CCAGAATAGNACTTGAAGCTGCATATCAACAAATTTAATCCAAAAACACCGATAACAAT 402  
 Db 113 ProArgValGluPheGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn 132  
 Qy 403 GATACTGATAATGGTGAATACTATAAACATTTTTCATTATCTCGTAAAGAT----- 453  
 Db 133 TyrLysAsnAspAla-----HisArgTyrCysAlaLeuGlyGlnGlnAspAsnSerGly 150  
 Qy 454 GCAATGGAAGATCAGCAATATGTAGTACTTAAAAAATGACGCGCATAACTTTTATGTCATTG 513  
 Db 151 IleProLysThrSerLysTyrValLeuLeuLysSerGluGlyLeuLeuAspIleSerPhe 170  
 Qy 514 ATGTTTAATCTCTGATGACATTACAGCTGAAGAGATATCTTTCGTACCATATGCGATGT 573  
 Db 171 MetLeuAsnAlaCysTyrAspIleIleAsnGluSerIleProLeuSerProTyrIleCys 190  
 Qy 574 GCAGGTATAGGACGACAGATCTTATCAGTATTTTAAAGACCTCAATCTAAAAATTTGCTTAC 633  
 Db 191 AlaGlyValGlyThrAspLeuIleSerMetPheGluAlaThrAsnProLysIleSerTyr 210  
 Qy 634 CAAGGAAAAATAGGTATTAGTTACCTTATCACACAGAAAGTCTCTGCAATTTATTGGTGA 693  
 Db 211 GlnGlyLysLeuGlyLeuSerTyrSerIleAsnProGluAlaSerValPheIleGlyGly 230  
 Qy 694 TACTACATGGCGTTATTGGTAAATAATTTGAGAAAGATACCTGTAATAACTCTCTGTAGTA 753  
 Db 231 HisPheHisLysValIleGlyAsnGluPheArgAspIleProThrLeuLysAlaPheVal 250  
 Qy 754 TTAATCATGCTCTCTCAACACCATCTGCTCACTAACTCTTGACGTTGGATACTTTGGC 813  
 Db 251 ThrSerSerAlaThrProAspLeu---AlaIleValThrLeuSerValCysHiePheGly 269

814 GGAGAAATTGGAATGAGGTTCACTTC 840  
 Db 270 IleGluLeuGlyGlyArgPheAsnPhe 278

RESULT 38  
 AAU73416  
 ID AAU73416 standard; Protein; 278 AA.  
 AC AAU73416;  
 DT 12-MAR-2002 (first entry)  
 XX Ehrlichia chaffeensis outer membrane protein p28-17.  
 DE Ehrlichia; outer membrane protein; p28; antibiotic; vaccine.  
 KW  
 KW Ehrlichia chaffeensis.  
 OS  
 XX  
 XX WO200183699-A2.  
 PN  
 XX  
 XX 08-NOV-2001.  
 PD  
 XX  
 XX 01-MAY-2001; 2001WO-US13997.  
 PF  
 XX  
 XX 01-MAY-2000; 2000US-201035P.  
 PR  
 XX  
 XX (RERE-) RES DEV FOUND.  
 PA  
 XX  
 XX Walker DH, Yu X;  
 PI  
 XX

DR WPI; 2002-066527/09.

XX Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated

PT P28 useful as a vaccine against Ehrlichia chaffeensis -

XX Disclosure; Figure 2; 97pp; English.

XX The invention relates to isolated and purified 28-kDa outer membrane

CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins

CC are encoded by a 28kDa outer membrane protein multigene family. P28

CC proteins are useful as a vaccine against E. chaffeensis. DNA encoding P28

CC is useful for transfecting a host cell. AAU73400-AAU73420 represent

CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.

XX

SQ Sequence 278 AA;

Alignment Scores:

Pred. No.:	1,596-56	Length:	278
Score:	577.00	Matches:	126
Percent Similarity:	60.5%	Conservative:	49
Best Local Similarity:	43.6%	Mismatches:	94
Query Match:	38.9%	Indels:	20
DB:	23	Gaps:	7

US-10-062-624-41 (1-840) x AAU73416 (1-278)

QY 1 ATGAATTATAGAAAATCTAGTAGAAGCGGTTAAATCTCAATTAATGTCAATCTTACCA 60

Db 1 MetAenCysLysPhePheleThrAlaLeuValSerLeuMetSerPheLeuPro 20

QY 61 TATCAGTCTTTGCGAGCTCTGAGTTCAGGTTCAAGAACTAATGATACAAAGAGGC---TTC 117

Db 21 GlyIleSerPheSerProVal-----GlnGlyAspAsnIleSerGlyAsnPhe 37

QY 118 TACATTAGTCAAGTACAAATCCAAAGTATATACACACTTTAGAAAATCTCTCTCAAGAA 177

Db 38 TyrValSerGlyTyrMetProSerAlaSerHisPheGlyMetSerAlaLysGlu 57

QY 178 ACTCCTATTATGGAACAAATCTCTCACTAAAGTTTCGGACTAAAGAAAGAT--- 234

Db 58 -----GluLysAsnProThrValAlaLeuTyrGlyLeuLysGlnAspTrp 72

QY 235 -----GGTGATATACAAAAGAGCAGATTTCACAGAGTAGCTCCAGGCATT 282

Db 73 GluGlyIleSerSerSerHisAsnAspAsnHisPheAsnAsnLysGlyTyrSerPhe 92

QY 283 GATTTCAAAATAACTTAATATCAGGATTTTCAGGAAGTATGTTACTCTATGACGCA 342

Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGlyGly 112

QY 343 CCAAGAAATAGAACTTGAGCTGCATATCAACAATTTATCCAAAACACCCGATACAAAT 402

Db 113 ProArgValGluPheGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn 132

QY 403 GATACGTGATAATGGTGAATCTATATAACATTTTGGATTATCTCGTAAAGAT----- 453

Db 133 TyrLysAsnAspAla-----HisArgTyrCysAlaLeuGlyGlnGlnAspAsnSerGly 150

QY 454 GCAATGGAAGATCAGCAATATAGTACTTAAATAATGACGGCATCACTTTATGTCAATG 513

Db 151 IleProLysThrSerLysTyrValLeuLeuLysSerGluGlyLeuLeuAspIleSerPhe 170

QY 514 ATGGTTAATCTGTTGATGACATTCAGCTGAGGAGATCTTTTCGTACCATCATGATGT 573

Db 171 MetLeuAsnAlaCysTyrAspIleIleAsnGluSerIleProLeuSerProTyrIleCys 190

QY 574 GCAGGTATAGGACGAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTAC 633

Db 191 AlaGlyValGlyThrAspLeuIleSerMetPheGluAlaThrAsnProLysIleSerTyr 210

QY 634 CAAGGAAAATAGGTATTAGTTACCTTATCACACCAAGAGTCTCTGCAATTTATGTTGGA 693

Db 211 GlnGlyLysLeuGlyLeuSerTyrSerIleAsnProGluAlaSerValPheIleGlyGly 230

QY 694 TACTACCATGCGGTTATGTTAAATTTGAGAGATACCTGTAAATAACTCTCTGACTA 753

Db 231 HisPheHisLysValIleGlyAsnGluPheArgAspIleProThrLeuLysAlaPheVal 250

QY 754 TTAATGATGCTCTCTCAAAACACATCTGCTTCACTAACTCTTTCAGTTGGATACCTTGGC 813

Db 251 ThrSerSerAlaThrProAspLeu--AlaIleValThrLeuSerValCysHisPheGly 269

QY 814 GGAGAAATGGAATGAGGTTCCACCTTC 840

Db 270 IleGluLeuGlyGlyArgPheAsnPhe 278

RESULT 39

AAU06942

ID AAY06942 standard; Protein; 256 AA.

XX AC AAY06942;

XX DT 05-JUL-1999 (first entry)

XX DE E. chaffeensis p28 protein.

XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;

XX detection; dog.

XX OS Ehrlichia chaffeensis.

XX PN WO9913720-A1.

XX PD 25-MAR-1999.

XX PF 18-SEP-1998; 98WO-US19600.

XX PR 19-SEP-1997; 97US-0059353.

XX PA (OHIS ) UNIV OHIO STATE.

XX PI Ohashi N, Rikihisa Y;

XX DR WPI; 1999-254290/21.

XX N-PSDB; AAX34742.

XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and

XX Ehrlichia canis

XX PS Claim 18; Fig 1; 55pp; English.

XX The invention provides isolated outer membrane proteins (OMP) from

XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part

XX of the OMP family and consist of proteins OMP-1, -1(B to Z) shown

XX in AAY06943-958. The E. canis proteins form part of the P30 family and

XX consist of proteins shown in AAY06959-970. The proteins and genes are

XX used to detect E. chaffeensis in patients and E. canis in dogs.

XX SQ Sequence 256 AA;

Alignment Scores:

Pred. No.:	1,856-55	Length:	256
Score:	567.50	Matches:	124
Percent Similarity:	60.53%	Conservative:	37
Best Local Similarity:	46.62%	Mismatches:	84
Query Match:	38.32%	Indels:	21
DB:	20	Gaps:	6

US-10-062-624-41 (1-840) x AAY06942 (1-256)

QY 76 GATCTGTAGCTCAAGACTAATCAACAAAGAGGCTTCTACATTAGTCAAGTAC 135

Db 1 AspProAlaGlySerGlyIleAsnGlyAsn-----PheTyrIleSerGlyLysTyr 17

QY 136 AATCCAAGTATATACACTTTAGAAAATTTCTGCTGAAGAACTCTTATTAATGAACA 195

Db 18 MetProSerAlaSerHisPheGlyValPheSerAlaLysGlu-----Glu 32  
 QY 196 AATCTCTCACTAAAAAGTTTCGGACTAAAGAAAGATGTGAT----- 240  
 Db 33 ArgAsnThrThrValGlyValPheGlyLeuLysGlnAenTrpAspGlySerAlaIleSer 52  
 QY 241 ---ATAACAAAAAGACGATTTCACAGAGTAGCTCCAGGCATGTATTTCAAAATAC 297  
 Db 53 AsnSerSerProAsnAspValPheThrValSerAsnTyrSerPheLysTyrGluAsnAsn 72  
 QY 298 TTAATATCAGGATTTTCAGGAAGATTGTTTACTCTATGACCGACCAAGAAATAGAACTT 357  
 Db 73 ProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetAspGlyProArgIleGluLeu 92  
 QY 358 GAAGCTGCATATCAACAAATTTAAATCCAAAAACACCGATAACAATGATGATGATGAT 417  
 Db 93 GluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn-----TyrLysAsn 110  
 QY 418 GAATACTATAACATTTTGCATTTATCTCGTAAAGATGCATGGAA-----GAT 465  
 Db 111 GluAlaHisArgTyrCysAlaLeuSerHisAsnSerAlaAlaAspMetSerSerAlaSer 130  
 QY 466 CAGCAATATGATGATCTAAAAAATGACGGCATAACTTTTATGTCATTGATGTTTAATACT 525  
 Db 131 AsnAsnPheValPheLeuLysAsnGluGlyLeuLeuAspIleSerPheMetLeuAsnAla 150  
 QY 526 TGCTATGACATTAACAGCTGAAGAGTATCTTTTCGTACCATATGATGTCAGGATATAGGA 585  
 Db 151 CysTyrAspValValGlyGluGlyIleProPheSerProTyrIleCysAlaGlyIleGly 170  
 QY 586 GCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAATA 645  
 Db 171 ThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSerTyrGlnGlyLysLeu 190  
 QY 646 GGTATTAGTACCCTATCACACAGAGTCTCTGATTATTGTTGGATACCTACCATGGC 705  
 Db 191 GlyLeuSerTyrSerIleSerProGluAlaSerValPheIleGlyGlyHisPheHisLys 210  
 QY 706 GTTATTGTAATAAATTTGAGAAGATACCTGTAATAACTCTCTGTAGTATTAAATGATGCT 765  
 Db 211 ValIleGlyAsnGluPheArgAspIleProThrIleIleProThrGlySerThrLeuAla 230  
 QY 766 CCTCAA---ACCACATCTGCTTCAGTAACTCTTACGTTGATGATCTTGGCGGAGAAAT 822  
 Db 231 GlyLysGlyAsnTyrProAlaIleValIleLeuAspValCysHisPheGlyIleGluLeu 250  
 QY 823 GGAATGAGGTTTCACCTTC 840  
 Db 251 GlyGlyArgPheValPhe 256

## RESULT 40

AAU73419  
 ID AAU73419 standard; Protein; 271 AA.

XX AC AAU73419;

XX DT 12-MAR-2002 (first entry)

XX DE Ehrlichia chaffeensis outer membrane protein P28-20.

XX KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.

XX OS Ehrlichia chaffeensis.

XX FN WO200183699-A2.

XX PD 08-NOV-2001.

XX PF 01-MAY-2001; 2001WO-US13997.

XX PR 01-MAY-2000; 2000US-201035P.

XX XX (RERE-) RES DEV FOUND.

XX Walker DH, Yu X;  
 PI WPI; 2002-066527/09.  
 DR Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated  
 XX P28 useful as a vaccine against Ehrlichia chaffeensis -  
 PT Claim 10; Figure 2; 97pp; English.  
 XX

CC The invention relates to isolated and purified 28-kDa outer membrane  
 CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins  
 CC are encoded by a 28kDa outer membrane protein multigene family. P28  
 CC proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28  
 CC is useful for transfecting a host cell. AAU73400-AAU73420 represent  
 CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.

SQ Sequence 271 AA;

Alignment Scores:  
 Pred. No.: 2,44e-55 Length: 271  
 Score: 566.50 Matches: 123  
 Percent Similarity: 61.35% Conservative: 50  
 Best Local Similarity: 43.62% Mismatches: 94  
 Query Match: 38.25% Indels: 15  
 DB: 23 Gaps: 9

US-10-062-624-41 (1-840) x AAU73419 (1-271)

QY 1 ATGATTTATAGAAATTTCTAGTAAGAGCGGTTATCTATCTATTAATGTCATCTTACCA 60  
 Db 1 MetAsnTyrLysLysPheValGlyValAlaLeuAlaThrLeuLeuSerPheLeuPro 20  
 QY 61 TATCAGCTTTTTCAGATCTCTGATGTTCAAGAACTAATGATCAACAAGAGGCTTCTAC 120  
 Db 21 AsnAsnSerPheSerAsp-----AlaAsnValProGluGlyArgLysGlyPheTyr 37  
 QY 121 ATTAGTCAAAAGTACAAATCCAAGTATATCACTTTAGAAAAATTTCTCTGCTGAAGAACT 180  
 Db 38 ValGlyThrGlnTyrLysValGlyValProAsnPheSerAsnPheSerAlaGluGluThr 57  
 QY 181 CCTATTAAATGCAACAAATTTCTCACTAAAAAGATTTCGACATAAAGAAAGAT---GGT 237  
 Db 58 -----LeuProGlyLeuThrLysSerIlePheAlaLeuGlyLeuAspLysSer 73  
 QY 238 GATATAACAAAAAAGACGATTTTACAAGAGTAGTCCAGGCATTTGATTTTCAAAATAAC 297  
 Db 74 SerIleSerAspHisAlaGlyPheThrGln---AlaTyrAsnProThrTyrAlaSerAsn 92  
 QY 298 TTAATATCAGGATTTTCAGGAAGTATTGTTTACTCTATGACCGACCAAGAAATAGAACTT 357  
 Db 93 ---PheAlaGlyPheGlyGlyValIleGlyTyrTyrValAsnAspPheArgValGluPhe 111  
 QY 358 GAAGCTGCATATCAACAAATTTAATCCAAAAACACCGATTAACATGATGATGATGATGAT 417  
 Db 112 GluGlyAlaTyrGluAsnPheGluProGluArg-----GlnTyrTyrProGluGlyGly 129  
 QY 418 GAATACTATAACATTTTGCATTTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTA 477  
 Db 130 GluSerHisLysPhePheAlaLeuSerArgGluSerThrValGlnAspAsnLysPheIle 149  
 QY 478 GTACTTTAAAAATGACCGCATTAACCTTTTATGTCATGATGTTTAAATATCTTCTATGACATT 537  
 Db 150 ValLeuGluAsnAspGlyValIleAspLysSerLeuAsnValAsnPheCysTyrAspIle 169  
 QY 538 ACAGCTGAAGGATATCTTTCGTACCATATGATGTCAGGATAGAGGATAGAGGATCTTATC 597  
 Db 170 AlaHisGlySerIleProLeuAlaProTyrMetCysAlaGlyValGlyAlaAspTyrIle 189  
 QY 598 ACTATTTTTAAACACCTCAATCTA---AAATTTGCTTACCAAGGAAAAATAGGTATTAGT 654  
 Db 190 Lys---PheLeuGlyIleSerLeuProLysPheSerTyrGlnValLysPheGlyValAsn 208



Db 81 AspGlyAspIleAlaGlnSerAlaAsnPheAsnArgThrAspProAlaLeuGluPheGln 100  
 QY 232 AATACCTAATATCAGGATTTTCAGGAGTATTGGTACTCTATGACGGACCAAGATA 351  
 Db 101 AsnAsnLeuIleSerGlyPheSerGlySerIleGlyTyrAlaMetAspGlyProargile 120  
 QY 352 GAACCTTGAGCTGCATATCAACATTTTAAATCCAAAAACACCGATAACAATGATCTGAT 411  
 Db 121 GluLeuGluAlaAlaTyrGlnLysPheAspAlaLysAsnProAspAsnAspThrAsn 140  
 QY 412 AATGCTGAATATATAAATCATTTTCATATCTCTGTAAGATGCAATGAAGATCAGCA 471  
 Db 141 SerGlyAspTyrTyrLysTyrPheGlyLeuSerArgGluAspAlaIleAlaAspLysLys 160  
 QY 472 TATGAGTACTTAAATACAGCGCATACCTTTTATGTCATTCATGCTTAAATCTGCTAT 531  
 Db 161 TyrValValLeuLysAsnGluGlyIleThrPheMetSerLeuMetValAsnThrCysTyr 180  
 QY 532 GACATTACAGCTGAAGGAGTATCTTTCGTACCATATGTCATGTCAGGTATAGGACAGAT 591  
 Db 181 AspIleThrAlaGluGlyValProPheIleProTyrAlaCysAlaGlyValGlyAlaAsp 200  
 QY 592 CTTATCCTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAATAGCTATT 551  
 Db 201 LeuIleAsnValPheLysAspPheAsnLeuLysPheSerTyrGlnGlyLysIleGlyIle 220  
 QY 652 ACTTACCCTTATCACACGAGTCTCTGCATTTATTTGGTGATACTTACCATGGCGTTATT 711  
 Db 221 SerTyrProIleThrProGluValSerAlaPheIleGlyGlyTyrThrHisGlyValIle 240  
 QY 712 GGTAATAAATTTGAGAAGATACCTGTAAATCACTCTGTAGTATTAAATGATGCTCTCAA 771  
 Db 241 GlyAsnAsnPheAsnLysIleProValIleThrProValValLeuGluGlyAlaProGln 260  
 QY 772 ACCACATCGCTTCAGTACTCTGAGCTGGATCTTGGCGGAGAAATTCGAATGAGG 831  
 Db 261 ThrThrSerAlaLeuValThrIleAspThrGlyTyrPheGlyGlyLeuValGlyValArg 280  
 QY 832 TTCACCTTC 840  
 Db 281 PheThrPhe 283

## RESULT 5

19-660-587-9  
 Sequence 9, Application US/09660587  
 Patent No. 6392023

## GENERAL INFORMATION:

APPLICANT: Walker, David H.

APPLICANT: McBride, Jere W.

APPLICANT: Yu, Xue-Jie

OP INVENTION: Homologous 28-kilodalton Immunodominant Protein

OP INVENTION: Genes of Ehrlichia canis and Uses Thereof

REFERENCE: D6152CIP2

APPLICATION NUMBER: US/09/660,587

FILING DATE: 2000-09-12

APPLICATION NUMBER: 09/261,358

FILING DATE: 1999-03-03

OP SEQ ID NOS: 46

NO 9

TH: 281

PR: PRT

NISM: Ehrlichia chaffeensis

URE:

ER INFORMATION: amino acid sequence of E. chaffeensis P28

60-587-9

## Sent Scores:

No.:	Length:	281
6.6e-66	Matches:	140
644.50	Conservative:	41
62.20%	Mismatches:	89
48.11%	Indels:	21
43.52%	Gaps:	6

## Similarity:

cal Similarity: 43.52%

ch: 4

US-10-062-624-41 (1-840) x US-09-660-587-9 (1-281)

QY 1 ATGAATATATAGAAAAATTTCTAGTAAGAGCGCGTAAATCTCTCAATTAATCTCAATCTTACCA 60  
 Db 1 MetAsnTyrLysLysValPheIleThrSerAlaLeuIleSerLeuIleSerSerLeuPro 20  
 QY 61 TATCAGCTCTTTTCAGAGATCTCTGTTAGGTTCAAGAACTAATATGATAACAAGAGGCTTCTAC 120  
 Db 21 GlyValSerPheSerAspProAlaGlySerGlyIleAsnGlyAsn-----PheTyr 37  
 QY 121 ATTAGTGCAAAGTACAACTCCAGATATATACACATTTAGAAAAATCTCTGCTGAAGAACT 180  
 Db 38 IleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu--- 56  
 QY 181 CCTATTATGCAACAAATTTCTCTCACTAAAAAGTTTTCGGACTAAAGAAAGATGGTGT 240  
 Db 57 -----GluArgAsnThrValGlyValPheGlyLeuLysGlnAsnTrpAsp 72  
 QY 241 -----ATACAAAAAGACGATTTTACAGAGTAGTCCAGGCATT 282  
 Db 73 GlySerAlaIleSerAsnSerSerProAsnAspValPheThrValSerAsnTyrSerPhe 92  
 QY 283 GATTTTCAAAATACTTAATATCAGGATTTTTCAGGAAGTATTGGTACTCTCTATGACGCA 342  
 Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetAspGly 112  
 QY 343 CCAGAATAGAACTTGAAGCTGCATATCAACAATTTTAAATCCAAAAACACCCGATCAACT 402  
 Db 113 ProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn 132  
 QY 403 GATACTGATAATGTTGAATATCTATAACATTTTGCATTTATCTCGTAAAGATGCAATGAA 462  
 Db 133 -----TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerHisAsnSerAlaAlaAsp 150  
 QY 463 -----GATCAGCAATATGTAAGTACTTAAATAATGACGCGATACCTTTTATGTCA 510  
 Db 151 MetSerSerAlaSerAsnAsnPheValPheLeuLysAsnGluGlyLeuLeuAspIleSer 170  
 QY 511 TTGATGGTTAATCTTGTCTGATGACATTAACAGCTGAAGGAGTATCTTCTGATCATATGCA 570  
 Db 171 PheMetLeuAsnAlaCysTyrAspValValGlyGluGlyIleProPheSerProTyrIle 190  
 QY 571 TGTGAGGTATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCT 630  
 Db 191 CysAlaGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSer 210  
 QY 631 TACCAAGGAAAAATAGGTATTAGTTACCTTATCACACCAAGAGTCTCTGCAATTTTGT 690  
 Db 211 TyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluLaserValPheIleGly 230  
 QY 691 GGATACTACCATGGCTTATTGGTAATAAATTTGAGAAGATACCTGTATAACTCTCTGTA 750  
 Db 231 GlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrIleIleProThr 250  
 QY 751 GTATTAAATGATGCTCTCAA---ACCACATCTGCTTCAGTAACTCTCTGACGTTGATAC 807  
 Db 251 GlySerThrLeuAlaGlyLysGlyAsnTyrProAlaIleValIleLeuAspValCysHis 270  
 QY 808 TTTGGCGGAGAAATTTGGAATGAGTTTCACCTTC 840  
 Db 271 PheGlyIleGluLeuGlyGlyArgPheAlaPhe 281

## RESULT 6

US-09-261-358A-9

Sequence 9, Application US/09261358A

Patent No. 6403780

GENERAL INFORMATION:

APPLICANT: Walker, David H.

APPLICANT: McBride, Jere W.

APPLICANT: Yu, Xue-Jie

OP INVENTION: Homologous 28-kilodalton Immunodominant Protein

OP INVENTION: Genes of Ehrlichia canis and Uses Thereof

281 PheThrPhe 283

f 3

261-358A-10  
nce 10, Application US/09261358A  
t No. 6403780

L INFORMATION:

CANT: Walker, David H.  
CANT: McBride, Jere W.  
CANT: Yu, Xue-Jie

OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
OF INVENTION: Genes of Ehrlichia canis and Uses Thereof

REFERENCE: D6152CIP

APPLICATION NUMBER: US/09/261,358A

FILING DATE: 1999-03-03

APPLICATION NUMBER: 09/201,458

FILING DATE: 1998-11-30

R OF SEQ ID NOS: 33

NO 10

CTH: 283

PE: PRT

ORGANISM: Ehrlichia chaffeensis

FEATURE:

OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1B

09-261-358A-10

Alignment Scores:

Pred. No.: 3,26e-130 Length: 283  
Score: 1202.50 Matches: 225  
Percent Similarity: 88.69% Conservative: 26  
Best Local Similarity: 79.51% Mismatches: 29  
Query Match: 81.20% Indels: 3  
DB: 4 Gaps: 1

US-10-062-624-41 (1-840) x US-09-261-358A-10 (1-283)

Qy 1 ATGAATTATAGAAATCTAGTAAGAGCGGTTAATCTCATTAATGCTCAATCTTACCA 60  
Db 1 MetAsnTyrLysLysPheValSerSerAlaLeuLeuSerLeuMetSerLeuPro 20  
Qy 61 TATCAGTCTTTGAGATCTCTGTAGTTCAAGA-----ACTAATGATAACAAGAA 111  
Db 21 TyrGlnSerPheAlaAaspProValThrSerAsnAaspThrGlyLeuAsnAaspArgGlu 40  
Qy 112 GGCTTCTACATTAGTCGAAAGTACATCCAGTATATCAGTTCAGTTCAGTTCCTGCT 171  
Db 41 GlyPheTyrLysSerValLysTyrAsnProSerLeuSerHisPheArgLysPheSerAla 60  
Qy 172 GAAGAACTCCTATTATTAATGGAACAAATCTCTCAATAAAAGTTTCGGACTAAAGAAA 231  
Db 61 GluGluAlaProLeuAsnGlyAsnThrSerLeuThrLysValPheGlyLeuLysLys 80  
Qy 232 GATGGTGATATAACAAGAGATCTTTTCAAGAGTAGTCTCCAGGCAATGATTTTCAA 291  
Db 81 AspGlyAspLeuAlaGlnSerAlaAsnPheAsnArgThrAspProAlaLeuGluPheGln 100  
Qy 292 AATACTTAATATCAGATTTTCAGGAGTATGTTACTCTATGCGGACGACCAAGATA 351  
Db 101 AsnAsnLeuLeuSerGlyPheSerGlySerLeuGlyTyrAlaMetAspGlyProArgGile 120  
Qy 352 GAATCTGAAGCTGATATCAAAATTTAATCCAAAACACCGATACCAATGATCTGAT 411  
Db 121 GluLeuGluAlaAlaTyrGlnLysPheAspAlaLysAsnProAspAsnAsnAaspThrAsn 140  
Qy 412 AATGGTGAATCTATAACATTTTGCATTATCTCGTAAAGATGCAATGGAAGATCAGCAA 471  
Db 141 SerGlyAspTyrTyrLysTyrPheGlyLeuSerArgGluAspAlaLeuAspLysLys 160  
Qy 472 TATGTACTTCTAAATGAGCGGATACCTTTTATGTCATGATGATGATGATGATGAT 531  
Db 161 TyrValValLeuLysAsnGluGlyLeuThrPheMetSerLeuMetValAsnThrCysTyr 180  
Qy 532 GACATTACAGCTGAAGAGTATCTTTCTGATACATATGATGTCAGGATATAGGACAGAT 591

Db 181 AspIleThrAlaGluGlyValProPheIleProTyrAlaCysAlaGlyValGlyAlaAasp 200  
Qy 592 CTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAATAGTATT 651  
Db 201 LeuIleAsnValPheLysAspPheAsnLeuLysPheSerTyrGlnGlyLysIleGlyLys 220  
Qy 652 AGTTACCTATCACACAGAAAGTCTCTGCATTTTATTTGTTGGATACTACCATGCGCTATT 711  
Db 221 SerTyrProIleThrProGluValSerAlaPheIleGlyGlyTyrTyrHisGlyValIle 240  
Qy 712 GGTAAATAATTTGAGAAGATACCTGTATATACTCTCTGTAGTATTAATGATGCTCTCAA 771  
Db 241 GlyAsnAsnPheAsnLysIleProValIleThrProValLeuLeuGluGlyAlaProGln 260  
Qy 772 ACCATATCTGCTTCAGTAACCTTTGACGTTGGATCTTTGGCGGAGAAATTTGGAATGAGG 831  
Db 261 ThrThrSerAlaLeuValThrIleAaspThrGlyTyrPheGlyGlyGluValGlyValArg 280  
Qy 832 TTCACCTTC 840  
Db 281 PheThrPhe 283

RESULT 4

US-09-201-458-6

Sequence 6, Application US/09201458A

Patent No. 6458942

GENERAL INFORMATION:

APPLICANT: Walker, David H.

APPLICANT: McBride, Jere W.

APPLICANT: Yu, Xue-Jie

TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia

TITLE OF INVENTION: canis and Uses Thereof

FILE REFERENCE: D6152

CURRENT APPLICATION NUMBER: US/09/201,458A

CURRENT FILING DATE: 1998-11-30

NUMBER OF SEQ ID NOS: 21

SEQ ID NO 6

LENGTH: 283

TYPE: PRT

ORGANISM: Ehrlichia chaffeensis

FEATURE:

OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1B

US-09-201-458-6

Alignment Scores:

Pred. No.: 3,26e-130 Length: 283  
Score: 1202.50 Matches: 225  
Percent Similarity: 88.69% Conservative: 26  
Best Local Similarity: 79.51% Mismatches: 29  
Query Match: 81.20% Indels: 3  
DB: 4 Gaps: 1

US-10-062-624-41 (1-840) x US-09-201-458-6 (1-283)

Qy 1 ATGAATTATAGAAATCTAGTAAGAGCGGTTAATCTCATTAATGCTCAATCTTACCA 60  
Db 1 MetAsnTyrLysLysPheValSerSerAlaLeuLeuSerLeuMetSerLeuPro 20  
Qy 61 TATCAGTCTTTGAGATCTCTGTAGTTCAAGA-----ACTAATGATAACAAGAA 111  
Db 21 TyrGlnSerPheAlaAaspProValThrSerAsnAaspThrGlyLeuAsnAaspArgGlu 40  
Qy 112 GGCTTCTACATTAGTCGAAAGTACATCCAGTATATCAGTTCAGTTCAGTTCCTGCT 171  
Db 41 GlyPheTyrLysSerValLysTyrAsnProSerLeuSerHisPheArgLysPheSerAla 60  
Qy 172 GAAGAACTCCTATTATTAATGGAACAAATCTCTCAATAAAAGTTTCGGACTAAAGAAA 231  
Db 61 GluGluAlaProLeuAsnGlyAsnThrSerLeuThrLysValPheGlyLeuLysLys 80  
Qy 232 GATGGTGATATAACAAGAGATCTTTTCAAGAGTAGTCTCCAGGCAATGATTTTCAA 291

1	MetAsnTyrLysLysIleLeuValArgSerAlaLeuIleSerLeuMetSerIleLeuPro	20
61	TATCAGTCTTTTGCAGATCTCTGTAGGTTCAAGAACTAATGATACAAAGAAGGCTTCTAC	120
21	TyrGlnSerPheAlaAspProValGlySerArgThrAsnAspAsnLysGluGlyPheTyr	40
121	ATTAGTGCAAAGTACAATCCAAGTATATACACATTTAGAAAAATCTCTGCTGAGAAACT	180
41	IleSerAlaLysTyrAsnProSerIleSerHisPheArgLysPheSerAlaGluGluThr	60
181	CGTATTATAGGAACAAATCTCTCACTAAAAAAGTTTTCGCACTAAAGAAAAGATGTTGAT	240
61	ProIleAsnGlyThrAsnSerLeuThrLysValPheGlyLeuLysLysAspGlyAsp	80
241	ATAACAAAAAAGACGATTTTCAAGTAGTAGCTCCAGGCATTGATTTTCAAATAACTT	300
81	IleThrLysLysAspAspPheThrArgValAlaProGlyIleAspPheGlnAsnLeu	100
301	ATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGACCGGACCAAGATAGACTTGAA	360
101	IleSerGlyPheSerGlySerIleGlyTyrSerMetAspGlyProArgIleGluLeuGlu	120
361	GCTGCATATCAACAAATTTAACTCCAAAAACACCGATAACAATCATCTCATGTAATGTGAA	420
121	AlaAlaTyrGlnGlnPheAsnProLysAsnThrAspAsnAspThrAspAsnGlyGlu	140
421	TACTATAACACATTTTGCATTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGATGA	480
141	TyrTyrLysHisPheAlaLeuSerArgLysAspAlaMetGluAspGlnTyrValVal	160
481	CTTAAAAATGACGGCATAACTTTTATGTCATTGATGTTTAACTTTCCTATGCAATACA	540
161	LeuLysAsnAspGlyIleThrPheMetSerLeuMetValAsnThrCysTyrAspIleThr	180
541	GCTGAAGGATCTTTTCGTACCATATGCAATGTCAGGTATAGAGCAGATCTTATCACT	600
181	AlaGluGlyValSerPheValProTyrAlaCysAlaGlyIleGlyAlaAspLeuIleThr	200
601	ATTTTTAAAGACCTCAATCTCAAAATTTGCTTTACCAAGAAAAATAGTATTAGTTACCT	660
201	IlePheLysAspLeuAsnLeuLysPheAlaTyrGlnGlyLysIleGlyIleSerTyrPro	220
661	ATCACACAGAGTCTCTGCATTTATTGGTGGATCTACCATGGCGTTATTGGTATATAA	720
221	IleThrProGluValSerAlaPheIleGlyTyrTyrHisGlyValIleGlyAsnLys	240
721	TTTGTAGAAGATACCTGTGTAATAACTCCTGTAGTATTAATGATGCTCTCAAACCAATCT	780
241	PheGluLysIleProValIleThrProValValLeuAsnAspAlaProGlnThrThrSer	260
781	GCTCAGTAACCTCTTGACGTTGGATACATTTTGGCGGAGAAATTGGATAGAGTTTACCTTC	840
261	AlaSerValThrLeuAspValGlyTyrPheGlyGlyIleGlyMetAlaPheThrPhe	280

## RESULT 2

US-09-660-587-10  
; Sequence 10, Application US/09660587  
; Patent No. 6392023  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; TITLE OF INVENTION: Genes of Ehrlichia canis and Usnes Thereof  
; FILE REFERENCE: D6152CIP2  
; RENT APPLICATION NUMBER: US/09/660,587  
; RENT FILING DATE: 2000-09-12  
; OR APPLICATION NUMBER: 09/261,358  
; OR FILING DATE: 1999-03-03  
; BER OF SEQ ID NOS: 46  
; NO 10  
; TH: 283  
; PRT

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

nucleic - protein search, using frame\_plus\_n2p model

Run on: July 8, 2003, 10:42:41 ; Search time 19 Seconds  
(without alignments)

2601.606 Million cell updates/sec

Title: US-10-062-624-41

Perfect score: 1481

Sequence: 1 atgaattataagaaattctt.....ttggaatgaggttcaccttc 840

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DSV=slp  
-Q=/cgn2\_1/usfto.spool/us10062624/runat\_08072003\_093520\_2056/app\_query.fasta\_1.1031  
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10062624 @CGN\_1\_1\_28 @runat\_08072003\_093520\_2056 -NCPU=6 -ICPU3  
-NO\_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA.\*

1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PTCUTS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1462	98.7	280	4	US-09-660-587-42
2	1202.5	81.2	283	4	Sequence 42, Appl
3	1202.5	81.2	283	4	Sequence 10, Appl
4	1202.5	81.2	283	4	US-09-660-587-10
5	644.5	43.5	281	4	US-09-201-458-6
6	644.5	43.5	281	4	US-09-660-587-9
7	644.5	43.5	281	4	US-09-261-358A-9
8	629.5	42.5	276	4	US-09-201-458-5
9	620	41.9	286	4	US-08-953-326-18
10	620	41.9	286	4	US-08-953-326-15
11	620	41.9	286	4	US-09-660-587-12
12	620	41.9	286	4	US-09-261-358A-12
13	620	41.9	286	4	US-09-201-458-8

13	605	40.9	280	4	US-09-660-587-14	Sequence 14, Appl
14	605	40.9	280	4	US-09-261-358A-14	Sequence 14, Appl
15	605	40.9	280	4	US-09-201-458-10	Sequence 10, Appl
16	603	40.7	280	4	US-08-953-326-17	Sequence 17, Appl
17	601	40.6	280	4	US-09-660-587-6	Sequence 6, Appl
18	601	40.6	280	4	US-09-261-358A-6	Sequence 6, Appl
19	582	39.3	278	4	US-09-660-587-2	Sequence 2, Appl
20	582	39.3	278	4	US-09-261-358A-2	Sequence 2, Appl
21	582	39.3	278	4	US-09-201-458-2	Sequence 2, Appl
22	577	39.0	278	4	US-08-953-326-16	Sequence 16, Appl
23	577	39.0	278	4	US-09-660-587-13	Sequence 13, Appl
24	577	39.0	278	4	US-09-261-358A-13	Sequence 13, Appl
25	577	39.0	278	4	US-09-201-458-9	Sequence 9, Appl
26	566	38.2	280	3	US-08-733-230-4	Sequence 4, Appl
27	566	38.2	280	4	US-08-953-326-4	Sequence 4, Appl
28	565	38.1	284	4	US-09-660-587-15	Sequence 15, Appl
29	565	38.1	284	4	US-09-261-358A-15	Sequence 15, Appl
30	565	38.1	284	4	US-09-201-458-11	Sequence 11, Appl
31	563	38.0	280	4	US-09-660-587-11	Sequence 11, Appl
32	563	38.0	280	4	US-09-261-358A-11	Sequence 11, Appl
33	563	38.0	280	4	US-09-201-458-7	Sequence 7, Appl
34	557.5	37.6	287	3	US-08-733-230-2	Sequence 2, Appl
35	557.5	37.6	287	4	US-08-953-326-2	Sequence 2, Appl
36	520	35.1	276	4	US-09-660-587-44	Sequence 44, Appl
37	515.5	34.8	287	4	US-08-953-326-19	Sequence 19, Appl
38	515.5	34.8	287	4	US-09-660-587-8	Sequence 8, Appl
39	515.5	34.8	287	4	US-09-261-358A-8	Sequence 8, Appl
40	515.5	34.8	287	4	US-09-201-458-4	Sequence 4, Appl
41	509.5	34.4	271	4	US-09-660-587-46	Sequence 46, Appl
42	509.5	34.4	283	4	US-09-660-587-4	Sequence 4, Appl
43	509.5	34.4	283	4	US-09-261-358A-4	Sequence 4, Appl
44	490.5	33.1	293	4	US-09-660-587-40	Sequence 40, Appl
45	386.5	26.1	282	3	US-08-733-230-6	Sequence 6, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-660-587-42  
; Sequence 42, Application US/09660587  
; Patent No. 6392023  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: US/09/660,587  
; CURRENT FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 42  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of E. canis p28-2 protein  
US-09-660-587-42

Alignment Scores:  
Pred. No.: 4,02e-160 Length: 280  
Score: 1462.00 Matches: 280  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 98.72% Indels: 0  
DB: 4 Gaps: 0

US-10-062-624-41 (1-840) x US-09-660-587-42 (1-280)

QY

1 ATGAATTATAAGAAATTCCTAGTAAGACGCGTTAATCTCAATATGTCATCTTACCA 60  
|||||



REFERENCE: D6152CIP  
 NENT APPLICATION NUMBER: US/09/261.358A  
 NENT FILING DATE: 1999-03-03  
 OR APPLICATION NUMBER: 09/201.458  
 OR FILING DATE: 1998-11-30  
 NUMBER OF SEQ ID NOS: 33  
 Q ID NO 9  
 LENGTH: 281  
 TYPE: PRT  
 ORGANISM: Ehrlichia chaffeensis

FEATURE:  
 OTHER INFORMATION: amino acid sequence of E. chaffeensis P28  
 US-09-261-358A-9

Alignment Scores:  
 Pred. No.: 281  
 Score: 6.6e-66 Length: 281  
 644.50 Matches: 140  
 Percent Similarity: 62.20% Conservative: 41  
 Best Local Similarity: 48.11% Mismatches: 89  
 Query Match: 43.52% Indels: 21  
 DB: 4 Gaps: 6

US-10-062-624-41 (1-840) x US-09-261-358A-9 (1-281)

```

QY 1 ATGAATTATAAGAAATTTCTAGTAAGAACGCGTTAAATCTCAATATGTCATCTTACCA 60
DB 1 MetAsnTyrLysLysValPheIleThrSerAlaLeuLeuSerLeuPro 20
QY 61 TATCAGTCTTTTGCAGATCTCTAGTTCAGAACTAATGATAACAAGAGCGTTCTAC 120
DB 21 GlyValSerPheSerAspProAlaGlySerGlyIleAsnGlyAsn-----PheTyr 37
QY 121 ATTAGTGCAGAACTAATCTCACTTATACACTTCTAGAAATTTCTGCTCAAGAACT 180
DB 38 IleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu--- 56
QY 181 CCTATTATGGAACAAATTTCTCTCACTTAAAGATTTTCGCACTAAAGAGATGGTAT 240
DB 57 -----GluArgAsnThrThrValGlyValPheGlyLeuLysGlnAsnTrpAsp 72
QY 241 -----ATACAAAAAGACGATTTTACAGAGTAGTCCAGGCATT 282
DB 73 GlySerAlaIleSerAsnSerSerProAsnAspValPheThrValSerAsnTyrSerPhe 92
QY 283 GATTTTCAAAATACTTAATATCAGGATTTTCAGAGAGTATTTGGTACTCTATGACCGA 342
DB 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyIleGlyTyrSerMetAspGly 112
QY 343 CCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGATACCAAT 402
DB 113 ProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn 132
QY 403 GATACTGATATGCGTGAATATATATAACATTTTCATTATCTCGTAAAGATGCAATGAA 462
DB 133 -----TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerHisAsnSerAlaAlaAsp 150
QY 463 -----GATCAGCAATATCTAGTACTTAAATAATGACGCGATACCTTTTATGCA 510
DB 151 MetSerSerAlaSerAsnAsnPheValPheLeuLysAsnGlyLeuLeuAspIleSer 170
QY 511 TTGATGGTTAATCTTGTATGACATTAACAGCTGAAGAGTATCTTCTGTCACATATGCA 570
DB 171 PheMetLeuAsnAlaCysTyrAspValValGlyGluGlyIleProPheSerProTyrIle 190
QY 571 TGTGAGGTATAGGAGAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCT 630
DB 191 CysAlaGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSer 210
QY 631 TACCAAGGAAAAATAGGTATTAGTACCTCATCACACAGAGTCTCTGCAATTTATGCT 690
DB 211 TyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSerValPheIleGly 230
QY 691 GGATACTACCTAGCGGTTATTGGTAAATAATTTGAGAGATACCTGTATAACTCCTGTA 750

```

```

DB 231 GlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrIleProThr 250
QY 751 GTATTAAATGATGCTCCTCAAA---ACCACATCTGCTCAGTAACTCTTACGCTTGATAC 807
DB 251 GlySerThrLeuAlaGlyLysGlyAsnTyrProAlaIleValIleLeuAspValCysHis 270
QY 808 TTTGCGGAGAAATTTGGAATGAGTTTCACCTTC 840
DB 271 PheGlyIleGluLeuGlyGlyArgPheAlaPhe 281

```

# RESULT 7

US-09-201-458-5  
 Sequence 5, Application US/09201458A  
 Patent No. 6458942

GENERAL INFORMATION:  
 APPLICANT: Walker, David H.  
 APPLICANT: McBride, Jere W.  
 APPLICANT: Yu, Xue-Jie  
 TITLE OF INVENTION: 28-kDa Immunoactive Protein Gene of Ehrlichia  
 TITLE OF INVENTION: canis and Uses Thereof  
 FILE REFERENCE: D6152  
 CURRENT APPLICATION NUMBER: US/09/201.458A  
 CURRENT FILING DATE: 1998-11-30  
 NUMBER OF SEQ ID NOS: 21  
 SEQ ID NO 5  
 LENGTH: 281  
 TYPE: PRT  
 ORGANISM: Ehrlichia chaffeensis  
 FEATURE:  
 OTHER INFORMATION: amino acid sequence of E. chaffeensis P28

US-09-201-458-5  
 Alignment Scores:  
 Pred. No.: 281  
 Score: 6.6e-66 Length: 281  
 644.50 Matches: 140  
 Percent Similarity: 62.20% Conservative: 41  
 Best Local Similarity: 48.11% Mismatches: 89  
 Query Match: 43.52% Indels: 21  
 DB: 4 Gaps: 6

US-10-062-624-41 (1-840) x US-09-201-458-5 (1-281)

```

QY 1 ATGAATTATAAGAAATTTCTAGTAAGAACGCGTTAAATCTCAATATGTCATCTTACCA 60
DB 1 MetAsnTyrLysLysValPheIleThrSerAlaLeuLeuSerLeuPro 20
QY 61 TATCAGTCTTTTGCAGATCTCTAGTTCAGAACTAATGATAACAAGAGCGTTCTAC 120
DB 21 GlyValSerPheSerAspProAlaGlySerGlyIleAsnGlyAsn-----PheTyr 37
QY 121 ATTAGTGCAGAACTAATCTCACTTATACACTTCTAGAAATTTCTGCTCAAGAACT 180
DB 38 IleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu--- 56
QY 181 CCTATTATGGAACAAATTTCTCTCACTTAAAGATTTTCGCACTAAAGAGATGGTAT 240
DB 57 -----GluArgAsnThrThrValGlyValPheGlyLeuLysGlnAsnTrpAsp 72
QY 241 -----ATACAAAAAGACGATTTTACAGAGTAGTCCAGGCATT 282
DB 73 GlySerAlaIleSerAsnSerSerProAsnAspValPheThrValSerAsnTyrSerPhe 92
QY 283 GATTTTCAAAATACTTAATATCAGGATTTTCAGAGAGTATTTGGTACTCTATGACCGA 342
DB 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyIleGlyTyrSerMetAspGly 112
QY 343 CCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGATACCAAT 402
DB 113 ProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn 132
QY 403 GATACTGATATGCGTGAATATATATAACATTTTCATTATCTCGTAAAGATGCAATGAA 462
DB 133 -----TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerHisAsnSerAlaAlaAsp 150
QY 463 -----GATCAGCAATATCTAGTACTTAAATAATGACGCGATACCTTTTATGCA 510
DB 151 MetSerSerAlaSerAsnAsnPheValPheLeuLysAsnGlyLeuLeuAspIleSer 170
QY 511 TTGATGGTTAATCTTGTATGACATTAACAGCTGAAGAGTATCTTCTGTCACATATGCA 570
DB 171 PheMetLeuAsnAlaCysTyrAspValValGlyGluGlyIleProPheSerProTyrIle 190
QY 571 TGTGAGGTATAGGAGAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCT 630
DB 191 CysAlaGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSer 210
QY 631 TACCAAGGAAAAATAGGTATTAGTACCTCATCACACAGAGTCTCTGCAATTTATGCT 690
DB 211 TyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSerValPheIleGly 230
QY 691 GGATACTACCTAGCGGTTATTGGTAAATAATTTGAGAGATACCTGTATAACTCCTGTA 750

```

Qy	61	TATCAGTCTTTTGCAGATCTCTGTAGGTTTCAAGAACTAATGATATCAACAAGAAAGGCTTCTAC	120
Db	21	GlyValSerPheSerAspProAlaGlySerGlyIleAsnGlyAsn-----PheTyr	37
Qy	121	ATTAGTCCAAGTACAATCCAGTATATCACACTTTAGAAAATTTCTCTGCTGAAGAAACT	180
Db	38	IleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu---	56
Qy	181	CCTATTAAATGGAACAAATCTCTCACTAAACAAAAGTTTTTCGACATAAAGAAAGATGGTGT	240
Db	57	-----GluArgAsnThrThrValGlyValPheGlyLeuLysGlnAsnTrpAsp	72
Qy	241	-----ATAACAAAAGACGATTTTCAAGAGTAGCTCCAGGCAT	282
Db	73	GlySerAlaIleSerAsnSerProAsnAspValPheThrValSerAsnTyrSerPhe	92
Qy	283	GATTTTCAAATAACTTAATATCAGGATTTTCAGGAAGTATGGTTACTCTATCGACGGA	342
Db	93	LysTyrGlnAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetAspGly	112
Qy	343	CCAAGATGAACCTTGAAGCTGTCATCAACAATTTTAATCCAAAAACACCGATAAACAAT	402
Db	113	ProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn	132
Qy	403	GATACTGATTAATCGTGAATACATAACAATTTTGCATTATCTCGTAAAGATGCAATGGAA	462
Db	133	-----TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerHisAsnSerAlaAlaAsp	150
Qy	463	-----GATCAGCAATATGATGACTTTAAAAATCAGCGCATTAATTTTATGTCA	510
Db	151	MetSerSerAlaSerAsnAsnPheValPheLeuLysAsnGluGlyLeuLeuAspIleSer	170
Qy	511	TTGATGGTTAACTTGTGATGACATACAGCTGAAGGAGTATCTTTTCGPACCATATGCA	570
Db	171	PheMetLeuAsnAlaCysTyrAspValValGlyGluGlyIleProPheSerProTyrIle	190
Qy	571	TGTGCAGGTATAGGACAGACTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCT	630
Db	191	CysAlaGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSer	210
Qy	631	TACCAAGGAAAATAGTATTAGTTACCTTATCACACCAAGAGCTCTGCACTATTATGGT	690
Db	211	TyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSerValPheIleGly	230
Qy	691	GGATACTACATGGCGTTATGGTAAATAATTGAGAAGATACCTGTAATACTCCTGTA	750
Db	231	GlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrIleIleProThr	250
Qy	751	GTATTAAATGATGCTCCTCTAC---ACCACATCTGCTTCAGTAACTCTTGACGTTGGATC	807
Db	251	GlySerThrLeuAlaGlyLysGlyAsnTyrProAlaIleValIleLeuAspValCysHis	270
Qy	808	TTTGGCGGAAATTTGGA	825
Db	271	PheGlyIleGluMetGly	276

b6

511 TTGATGGTTAACTACTTCCTATGACATTACAGCTGAAGGAGTATCTTTCGTACCATATGCA 570  
171 PheMetLeuAsnAlaCysTyrAspValValGlyGluGlyIleProPheSerProTyrIle 190  
571 TGTGCAGGTATAGGACAGACTCTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCT 630  
191 CysAlaGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSer 210  
631 TACCAAGGAAAATAGTATTAGTTACCTATCACAACCGAAGCTCTCTGCATTTATTTGGT 690  
211 TyrGlnGlyLysLeuGlyLeuGlySerTyrSerIleSerProGluAlaSerValPheIleGly 230  
691 GGATACTACCATGGCGCTTATTGGTATAATAATTGGAGAAGACTCTGTAATACTCCTGTGA 750  
231 GlyHisPheHisLysValleGlyAsnGluPheArgAspIleProThrIleIleProThr 250  
751 GTATTAAATGATGCTCCTCAA---ACCACATCTGCTTCAGTAACTCTTGACGTTGGATAC 807  
251 GlySerThrLeuAlaGlyLysGlyAsnTyrProAlaIleValIleLeuAspValCysHis 270  
808 TTGGCGGAGAAATTTGGA 825  
271 PheGlyIleGluMetGly 276

RESULT 9  
US-08-953-326-15  
; Sequence 15, Application US/08953326  
; Patent No. 6251872  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony P.  
; APPLICANT: Ganta, Roman R.  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Wika, Aceme  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Mahan, Suman M.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of  
; TITLE OF INVENTION: Animals and Humans  
; TITLE REFERENCE: UF-167C1  
; CURRENT APPLICATION NUMBER: US/08/953,326  
; FILING DATE: 1997-10-17

```

: Sequence 15, Application US/08953326
: Patent No. 6251872
: GENERAL INFORMATION:
: APPLICANT: Barbet, Anthony F.
: APPLICANT: Ganta, Roman R.
: APPLICANT: McGuire, Travis C.
: APPLICANT: Burridge, Michael J.
: APPLICANT: Wyika, Aceme
: APPLICANT: Rurangirwa, Fred R.
: APPLICANT: Mahan, Suman M.
: TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
: TITLE OF INVENTION: Animals and Humans
: FILE REFERENCE: UF-167C1
: CURRENT APPLICATION NUMBER: US/08/953,326
: FILING DATE: 1997-10-17

```

```
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-15

Alignment Scores:
Pred. No.: 4,43e-63 Length: 286
Score: 620.00 Matches: 139
Percent Similarity: 61.56% Conservative: 42
Best Local Similarity: 47.28% Mismatches: 89
Query Match: 41.86% Indels: 24
DB: 4 Gaps: 7

US-10-062-624-41 (1-840) x US-08-953-326-15 (1-286)
Qy 1 ATGAATTATAAGAAATCTAGTAAGAGCGGTTAATCTCATTAAATGCTCAATCTTACCA 60
Db 1 MetAenCysGluLysPhePheleThrAlaLeuThrLeuLeuMetSerPheLeuPro 20
Qy 61 TATCAGTCTTTTGCAGATCTGTAGTTTCAAGAACTAATGATAACAAAGAGGC---TTC 117
Db 21 GlyileSerLeuSerAspProVal-----GlnAspAspAsnIleSerGlyAsnPhe 37
Qy 118 TACATTAGTCAAGTACAAATCCAAAGTATATACACATTTAGAAAAATCTCTGCTGAAGAA 177
Db 38 TyriLeSerGlyLysTyMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu 57
Qy 178 ACTCCTATTAAATGAACAAATCTCTCACTAAAAAGTTTTCGAGCTAAAGAAAGATGGT 237
Db 58 -----GluArgAsnThrThrValGlyValPheGlyIleGluGlnAspTrp 72
Qy 238 GAT-----ATAACAAAAAGACGATTTTACAGAGTAGTCCAGGC 279
Db 73 AspArgCysValIleSerArgThrThrLeuSerAspIlePheThrValProAsnTySer 92
Qy 280 ATTGATTTTCAAAATACCTAATATCAGGATTTTCAGGAAGTATGGTTACTCTATGAC 339
Db 93 PheLysTyGlnGlyIleGlyLeuAspLeuValSerMetPheGluAlaIleAsnPro 210
Qy 622 AAATTTGTTACCAAGGAAATAGGTATTAGTTACCTATCACACGAGAGTCTCTGCA 681
Db 211 LysileSerTyGlnGlyLysLeuGlyLeuSerTyProIleSerProGluAlaSerVal 230
Qy 682 TTTATGGTGGTACTACCATGGGTTTATGGTAAATTAATTTGAGAGATACCTGTATA 741
Db 231 PheileGlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrMet 250

US-10-062-624-41 (1-840) x US-09-660-587-12 (1-286)
Qy 1 ATGAATTATAAGAAATCTAGTAAGAGCGGTTAATCTCATTAAATGCTCAATCTTACCA 60
Db 1 MetAenCysGluLysPhePheleThrAlaLeuThrLeuLeuMetSerPheLeuPro 20
Qy 61 TATCAGTCTTTTGCAGATCTGTAGTTTCAAGAACTAATGATAACAAAGAGGC---TTC 117
Db 21 GlyileSerLeuSerAspProVal-----GlnAspAspAsnIleSerGlyAsnPhe 37
Qy 118 TACATTAGTCAAGTACAAATCCAAAGTATATACACATTTAGAAAAATCTCTGCTGAAGAA 177
Db 38 TyriLeSerGlyLysTyMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu 57
Qy 178 ACTCCTATTAAATGAACAAATCTCTCACTAAAAAGTTTTCGAGCTAAAGAAAGATGGT 237
Db 58 -----GluArgAsnThrThrValGlyValPheGlyIleGluGlnAspTrp 72
Qy 238 GAT-----ATAACAAAAAGACGATTTTACAGAGTAGTCCAGGC 279
Db 73 AspArgCysValIleSerArgThrThrLeuSerAspIlePheThrValProAsnTySer 92
Qy 280 ATTGATTTTCAAAATACCTAATATCAGGATTTTCAGGAAGTATGGTTACTCTATGAC 339
Db 93 PheLysTyGlnGlyIleGlyLeuAspLeuValSerMetPheGluAlaIleAsnPro 210
Qy 622 AAATTTGTTACCAAGGAAATAGGTATTAGTTACCTATCACACGAGAGTCTCTGCA 681
Db 211 LysileSerTyGlnGlyLysLeuGlyLeuSerTyProIleSerProGluAlaSerVal 230
Qy 682 TTTATGGTGGTACTACCATGGGTTTATGGTAAATTAATTTGAGAGATACCTGTATA 741
Db 231 PheileGlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrMet 250

Alignment Scores:
Pred. No.: 4,43e-63 Length: 286
Score: 620.00 Matches: 139
Percent Similarity: 61.56% Conservative: 42
Best Local Similarity: 47.28% Mismatches: 89
Query Match: 41.86% Indels: 24
DB: 4 Gaps: 7

US-10-062-624-41 (1-840) x US-09-660-587-12 (1-286)
Qy 1 ATGAATTATAAGAAATCTAGTAAGAGCGGTTAATCTCATTAAATGCTCAATCTTACCA 60
Db 1 MetAenCysGluLysPhePheleThrAlaLeuThrLeuLeuMetSerPheLeuPro 20
Qy 61 TATCAGTCTTTTGCAGATCTGTAGTTTCAAGAACTAATGATAACAAAGAGGC---TTC 117
Db 21 GlyileSerLeuSerAspProVal-----GlnAspAspAsnIleSerGlyAsnPhe 37
Qy 118 TACATTAGTCAAGTACAAATCCAAAGTATATACACATTTAGAAAAATCTCTGCTGAAGAA 177
Db 38 TyriLeSerGlyLysTyMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu 57
Qy 178 ACTCCTATTAAATGAACAAATCTCTCACTAAAAAGTTTTCGAGCTAAAGAAAGATGGT 237
Db 58 -----GluArgAsnThrThrValGlyValPheGlyIleGluGlnAspTrp 72
Qy 238 GAT-----ATAACAAAAAGACGATTTTACAGAGTAGTCCAGGC 279
Db 73 AspArgCysValIleSerArgThrThrLeuSerAspIlePheThrValProAsnTySer 92
Qy 280 ATTGATTTTCAAAATACCTAATATCAGGATTTTCAGGAAGTATGGTTACTCTATGAC 339
Db 93 PheLysTyGlnGlyIleGlyLeuAspLeuValSerMetPheGluAlaIleAsnPro 210
Qy 622 AAATTTGTTACCAAGGAAATAGGTATTAGTTACCTATCACACGAGAGTCTCTGCA 681
Db 211 LysileSerTyGlnGlyLysLeuGlyLeuSerTyProIleSerProGluAlaSerVal 230
Qy 682 TTTATGGTGGTACTACCATGGGTTTATGGTAAATTAATTTGAGAGATACCTGTATA 741
Db 231 PheileGlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrMet 250
```



Alignment Scores:

Pred. No.: 4,43e-63 Length: 286  
Score: 620.00 Matches: 139  
Percent Similarity: 61.56% Conservative: 42  
Best Local Similarity: 47.28% Mismatches: 89  
Query Match: 41.86% Indels: 24  
DB: 4 Gaps: 7

US-10-062-624-41 (1-840) x US-09-201-458-8 (1-286)

```
Qy 1 ATGAATTATAAGAAATTCCTAGTAAGAGCGCGTTAAATCTCATTAATGCTCAATCTTACCA 60
Db 1 MetAsnCySylsPhePheIleThrThrAlaLeuThrLeuLeuMetSerPheLeuPro 20
Qy 61 TATCAGTCTTTTGCAGATCTGTAGTTTCAAGAACTAATGATAACAAGAGGC---TTC 117
Db 21 GlyIleSerLeuSerAspProVal-----GlnAspAspAsnIleSerGlyAsnPhe 37
Qy 118 TACATTAGTCAAGTACAATCCAAAGTATATCACATTTAGAAAATTTCTGCTGGAAGAA 177
Db 38 TyrIleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu 57
Qy 178 ACTCTATTATGAACAATTTCTCACTAAAGTTTTCGGACTAAAGAAAGAGTGT 237
Db 58 -----GluArgAsnThrThrValGlyValPheGlyIleGlnAspTrp 72
Qy 238 GAT-----ATAACAAAAGAGCGATTTTACAAGAGTAGTCCAGGC 279
Db 73 AspArgCysValIleSerArgThrThrLeuSerAspIlePheThrValProAsnTyrSer 92
Qy 280 ATTGATTTTCAAAATACTTAATATCAGGATTTTCAGGAAGTATGTTGTTACTATGGAAC 339
Db 93 PheLysTyrGluAsnAsnLeuPheSerGlyPheAlaGlyAlaIleGlyTyrSerMetAsp 112
Qy 340 GGACCAAGATAGAACTTGAAGTCCATCAACAAATTTAATCCAAAACACCGATAAC 399
Db 113 GlyProArgIleGluLeuGluValSerTyrGluAlaPheAspValLysAsnGlnGlyAsn 132
Qy 400 AATGATACATAATGTTGAATACATAAACAATTTTGCATTATCTCGTAAAGATGCAATG 459
Db 133 Asn-----TyrLysAsnGluAlaHisArgTyrTyrAlaLeuSerHisLeuLeuGlyThr 150
Qy 460 GAAGATCAG-----CAATATGTAGTACTTAAATAATGACGGCATAACT 501
Db 151 GluThrGlnIleAspGlyAlaGlySerAlaSerValPheLeuIleAsnGluGlyLeuLeu 170
Qy 502 TTTATGTCATGTGGTTAAATCTGCTATGACATTACAGTCTGAGGAGTATCTTTCGTA 561
Db 171 AspLysSerPheMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSer 190
Qy 562 CCATATGTCATGTGAGGATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTA 621
Db 191 ProTyrIleCysAlaGlyIleGlyIleAspLeuValSerMetPheGluAlaIleAsnPro 210
Qy 622 AAATTTGCTTACCAAGGAAAATAGTATTAGTACCCCTATCACACCAAGAGTCTCTGCA 681
Db 211 LysIleSerTyrGlnGlyLysLeuGlyLeuSerTyrProIleSerProGluAlaSerVal 230
Qy 682 TTTATTTGGTGATACCTACCGGCTTATGCTAATAATTTGAGAAGATACCTGTAATA 741
Db 231 PheIleGlyIleHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrMet 250
Qy 742 ACTCCTGTAGTATTAAATGATGCTCCTCAA---ACCACATCTGCTTCAAGTAACTCTTGAC 798
Db 251 IleProSerGluSerAlaLeuAlaGlyLysGlyAsnTyrProAlaIleValThrLeuAsp 270
Qy 799 GTTGGATCTTTGGCGGAGAAATTTGGAATGAGTTACCTTCACCTTC 840
Db 271 ValPheTyrPheGlyIleGluLeuGlyGlyArgPheAsnPhe 284
```

RESULT 13

US-09-660-587-14

; Sequence 14, Application US/0960587

Patent No. 6392023

GENERAL INFORMATION:  
APPLICANT: Walker, David H.  
APPLICANT: McBride, Jere W.  
APPLICANT: Yu, Xue-Jie  
TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof  
FILE REFERENCE: D6152CIP2  
CURRENT APPLICATION NUMBER: US/09/660,587  
CURRENT FILING DATE: 2000-09-12  
PRIOR APPLICATION NUMBER: 09/261,358  
PRIOR FILING DATE: 1999-03-03  
NUMBER OF SEQ ID NOS: 46  
SEQ ID NO 14

LENGTH: 280  
TYPE: PRT  
ORGANISM: Ehrlichia chaffeensis  
FEATURE:

OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F  
US-09-660-587-14

Alignment Scores:

Pred. No.: 2,35e-61 Length: 280  
Score: 605.00 Matches: 133  
Percent Similarity: 63.39% Conservative: 54  
Best Local Similarity: 45.08% Mismatches: 78  
Query Match: 40.85% Indels: 30  
DB: 4 Gaps: 10

US-10-062-624-41 (1-840) x US-09-660-587-14 (1-280)

```
Qy 1 ATGAATTATAAGAAATTCCTAGTAAGAGCGCGTTAAATCTCATTAATGCTCAATCTTACCA 60
Db 1 MetAsnCySylsPhePheIleThrThrLeuValSerLeuMetSerPheLeuPro 20
Qy 61 TATCAGTCTTTTGCAGATCTGTAGTTTCAAGAACTAATGATAACAAGAGGC---TTC 117
Db 21 GlyIleSerPheSerAspAlaVal-----GlnAsnAspAsnValGlyGlyAsnPhe 37
Qy 118 TACATTAGTCAAGTACAATCCAAAGTATATCACATTTAGAAAATTTCTGCTGGAAGAA 177
Db 38 TyrIleSerGlyLysTyrValProSerValSerHisPheGlyValPheSerAlaLysGln 57
Qy 178 ACTCCTATTATGAACAATTTCTCTCACTAAAGAGTCTTTCGGACTAAAGAAAGAGTGT 237
Db 58 -----GluArgAsnThrThrGlyValPheGlyLeuLysGlnAspTrp 72
Qy 238 GAT-----ATAACAAAAGAGC---GATTTTACAAGAGTAGTCCAGGATTCAT 285
Db 73 AspGlySerThrIleSerLysAsnSerProGluAsnThrPheAsnValProAsnTyrSer 92
Qy 286 TTT-----CAAAATACTTAATATCAGGATTTTCAGGAAGTATGCTTACTCTATCGAC 339
Db 93 PheLysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaValGlyTyrLeuMetAsn 112
Qy 340 GGACCAAGATAGAACTTGAAGTCCATATCAACAAATTTAATCCAAAACACCGATAAC 399
Db 113 GlyProArgIleGluLeuGluMetSerTyrGluThrPheAspValLysAsnGlnGlyAsn 132
Qy 400 AATGATACATAATGTTGAATACATAAACAATTTTGCATTATCTCGT----- 447
Db 133 AsnTyrLysAsnAspAla-----HisLysTyrTyrAlaLeuThrHisAsnSerGlyGly 150
Qy 448 AAAGATCCAATGGAAGATCAGCAATATGATAGTCTTAAATAATGACGCATATCTTATG 507
Db 151 LysLeuSerAsnAlaGlyAspLysPheValPheLeuLysAsnGluGlyLeuLeuAspIle 170
Qy 508 TCATTGATGTTAATATCTGCTATGACATTACAGTCAAGAGGATGATCTTTCGTAACCAT 567
Db 171 SerLeuMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSerProTyr 190
Qy 568 GCATGTCAGGTATAGGAGAGATCTTATCACTATTTTAAAGACCTCAATCTTAAATTT 627
Db 568 GCATGTCAGGTATAGGAGAGATCTTATCACTATTTTAAAGACCTCAATCTTAAATTT
```

Db 191 IleCysAlaGlyValGlyThrAspLeuIleSerMetPheGluAlaLeuAsnProLysile 210  
Qy 628 GCTTACCAAGAGAAATAGGTATTAGTACCTATCACACCAAGAGTCTCTGCATTATT 687  
Db 211 SerTyGlnGlyLeuGlyLeuSerTySerIleSerProGluAlaSerValPheVal 230  
Qy 688 GGTGATACCTACCATGGCGTATTGGTAATAATTTGAGAGATACCTGTAATACTCCT 747  
Db 231 GlyGlyHisPheHisLeuValIleGlyAsnGluPheArgAspIleProAlaMetIlePro 250  
Qy 748 GTAGTATTAAATGATGCTCCTCAACACCATCTGCTTCA-----GTAACCTTT 795  
Db 251 -----SerThrSerThrLeuThrGlyAsnHisPheThrIleValThrLeu 265  
Qy 796 GACGTTGATACCTTTGGCGGAGAAATGGAAATGAGGTTACACCTTC 840  
Db 266 SerValCysHisPheGlyValGluLeuGlyGlyArgPheAsnPhe 280  
RESULT 14  
US-09-261-358A-14  
; Sequence 14, Application US/09261358A  
; Patent No. 6403780  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP  
; CURRENT APPLICATION NUMBER: US/09/261,358A  
; CURRENT FILING DATE: 1999-03-03  
; PRIOR APPLICATION NUMBER: 09/201,458  
; PRIOR FILING DATE: 1998-11-30  
; NUMBER OF SEQ ID NOS: 33  
; SEQ ID NO 14  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis.  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F  
US-09-261-358A-14  
Alignment Scores:  
Pred. No.: 2,358-61 Length: 280  
Score: 605.00 Matches: 133  
Percent Similarity: 63.39% Conservative: 54  
Best Local Similarity: 45.08% Mismatches: 78  
Query Match: 40.85% Indels: 30  
DB: 4 Gaps: 10  
US-10-062-624-41 (1-840) x US-09-261-358A-14 (1-280)  
Qy 1 ATGAATTATAAGAAATCTAGTAAGAGCGGTAAATCTCAATTAATGTCAATCTTACCA 60  
Db 1 MetAsnCysValGlyPhePheIleThrThrLeuValSerLeuMetSerPheLeuPro 20  
Qy 61 TATAGTCTTTGGCAGATCTGTAGTTCAGAACTAATGATAACAAAGAGCG---TTC 117  
Db 21 GlyIleSerPheSerAspAlaVal-----GlnAsnAspAsnValGlyGlyAsnPhe 37  
Qy 118 TACATTAGTCAAGATGACAACTCAATATATACACTTTAGAAAATCTCTGCTGAAGAA 177  
Db 38 TyrlleSerGlyLysTyValProSerValSerHisPheGlyValPheSerAlaLysGln 57  
Qy 178 ACTCCTATTATGAACAAATCTCTCACTAAAAAGTTTCGGACTAAAGAAAGATGGT 237  
Db 58 -----GluArgAsnThrThrGlyValPheGlyLeuLysGlnAspTrp 72  
Qy 238 GAT-----ATACAAAAAAGAC-----GATTTTACAGAGTAGTCCAGGCATTTGAT 285  
Db 73 AspGlySerThrIleSerLysAsnSerProGluAsnThrPheAsnValProAsnTySer 92  
Qy 286 TTT-----CAAATAACTTAATATATACAGGATTTTCAGGAAGATTGGTTACTCTATGAC 339

Db 93 PheLysTyGluAsnAsnProPheLeuGlyPheAlaValGlyTyLeuMetAsn 112  
Qy 340 GGACCAAGAATAGAACTTGAAGCTGCATATCAACAATTTTAAATCCAAAAACACCGATAAC 399  
Db 113 GlyProArgIleGluLeuGluMetSerTyGluThrPheAspValLysAsnGlnGlyAsn 132  
Qy 400 AATGATACTGATATGTTGAATACATATAAATTTTGCATTATCTCGT----- 447  
Db 133 AsnTyrlsAsnAspAla-----HisLysTyTyAlaLeuThrHisAsnSerGlyGly 150  
Qy 448 AAAGATGCAATGGAAGATCAGCAATATGTACTTAAATAATGACGGCATAACTTTTATG 507  
Db 151 LysLeuSerAsnAlaGlyAspLysPheValPheLeuLysAsnGluGlyLeuAspIle 170  
Qy 508 TCATTGATGGTTAATCTGCTATGACATTACAGCTCAAGAGAGTATCTTCGTACCATAT 567  
Db 171 SerLeuMetLeuAsnAlaCysTyAspValIleSerGluGlyIleProPheSerProTy 190  
Qy 568 GCATGTCAGGTATAGGAGCATCTTATCACTATTTTAAAGACCTCAATCTAAATTT 627  
Db 191 IleCysAlaGlyValGlyThrAspLeuIleSerMetPheGluAlaIleAsnProLysile 210  
Qy 628 GCTTACCAAGAGAAATAGGTATTAGTTACCTATCACACCAAGTCTCTGCAATTTATT 687  
Db 211 SerTyGlnGlyLysLeuGlyLeuSerTySerIleSerProGluAlaSerValPheVal 230  
Qy 688 GGTGATACCTACCATGGCGTATTGGTAATAATTTGAGAGATACCTGTAATACTCCT 747  
Db 231 GlyGlyHisPheHisLeuValIleGlyAsnGluPheArgAspIleProAlaMetIlePro 250  
Qy 748 GTAGTATTAAATGATGCTCCTCAACACCATCTCTCTCA-----GTAACCTTT 795  
Db 251 -----SerThrSerThrLeuThrGlyAsnHisPheThrIleValThrLeu 265  
Qy 796 GACGTTGATACCTTTGGCGGAGAAATGGAAATGAGGTTACACCTTC 840  
Db 266 SerValCysHisPheGlyValGluLeuGlyGlyArgPheAsnPhe 280  
RESULT 15  
US-09-201-458-10  
; Sequence 10, Application US/09201458A  
; Patent No. 6458942  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia  
; FILE REFERENCE: D6152  
; CURRENT APPLICATION NUMBER: US/09/201,458A  
; CURRENT FILING DATE: 1998-11-30  
; NUMBER OF SEQ ID NOS: 21  
; SEQ ID NO 10  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F  
US-09-201-458-10  
Alignment Scores:  
Pred. No.: 2,358-61 Length: 280  
Score: 605.00 Matches: 133  
Percent Similarity: 63.39% Conservative: 54  
Best Local Similarity: 45.08% Mismatches: 78  
Query Match: 40.85% Indels: 30  
DB: 4 Gaps: 10  
US-10-062-624-41 (1-840) x US-09-201-458-10 (1-280)  
Qy 1 ATGAATTATAAGAAATCTAGTAAGAGCGGTAAATCTCAATTAATGTCAATCTTACCA 60  
Db 1 MetAsnCysValGlyPhePheIleThrThrLeuValSerLeuMetSerPheLeuPro 20  
Qy 61 TATAGTCTTTGGCAGATCTGTAGTTCAGAACTAATGATAACAAAGAGCG---TTC 117  
Db 21 GlyIleSerPheSerAspAlaVal-----GlnAsnAspAsnValGlyGlyAsnPhe 37  
Qy 118 TACATTAGTCAAGATGACAACTCAATATATACACTTTAGAAAATCTCTGCTGAAGAA 177  
Db 38 TyrlleSerGlyLysTyValProSerValSerHisPheGlyValPheSerAlaLysGln 57  
Qy 178 ACTCCTATTATGAACAAATCTCTCACTAAAAAGTTTCGGACTAAAGAAAGATGGT 237  
Db 58 -----GluArgAsnThrThrGlyValPheGlyLeuLysGlnAspTrp 72  
Qy 238 GAT-----ATACAAAAAAGAC-----GATTTTACAGAGTAGTCCAGGCATTTGAT 285  
Db 73 AspGlySerThrIleSerLysAsnSerProGluAsnThrPheAsnValProAsnTySer 92  
Qy 286 TTT-----CAAATAACTTAATATATACAGGATTTTCAGGAAGATTGGTTACTCTATGAC 339

Db 1 MetAenCysLysLysPhePheIleThrThrLeuValSerLeuMetSerPheLeuPro 20  
Qy 61 TATCAGTCTTTTGCAGATCTCTAGGTTCAAGAACTAATGATAACAAAGAAGGC---TTC 117  
Db 21 GlyIleSerPheSerAlaVal-----GlnAenAenValGlyGlyAsnPhe 37  
Qy 118 TACATTAGTCAAGTACAATCAAGTATATACACTTTAGAAAATTTCTCTGCTGAAGAA 177  
Db 38 TyrIleSerGlyLysTyrValProSerValSerHisPheGlyValPheSerAlaLysGln 57  
Qy 178 ACTCCTATTAAATGGAACAAATCTCTCACTAAAGTTTCGGAGTAAAGAAAGATGGT 237  
Db 58 -----GluArgAsnThrThrGlyValPheGlyLeuLysGlnAspTrp 72  
Qy 238 GAT-----ATAACAAAAAGAC---GATTTTACAAGAGTAGCTCCAGCGCATTTGAT 285  
Db 73 AspGlySerThrIleSerLysAenSerProGluAenThrPheAenValProAenTyrSer 92  
Qy 286 TTT-----CAAAATAACTTAATCAGATTTTCAGGAAGTATTTGTTACTCTATGGAC 339  
Db 93 PheLysTyrGluAenAenProPheLeuGlyPheAlaGlyAlaValGlyTyrLeuMetAen 112  
Qy 340 GGACCAAGATAGACTTGAAGTCTCATATCAACAATTTTAAAGACCTCAATCTAAATTT 399  
Db 113 GlyProArgIleGluLeuGluMetSerTyrGluThrPheAenValLysAenGlnGlyAen 132  
Qy 400 AATGATCTGATTAATGAGTCTCTCAACCAATTTTAAAGACCTCAATCTAAATTT 447  
Db 133 AsnTyrLysAenAenAenAenAenAenAenAenAenAenAenAenAenAenAenAen 150  
Qy 448 AAGATGCAATGGAAGATCAGCAATATAGTACTTAAATATGAGCGGATTAATTTATG 507  
Db 151 LysLeuSerAenAlaGlyAspLysPheValPheLeuLysAenGluGlyLeuAenAspIle 170  
Qy 508 TCATTGATGGTTAATCTTGTATGACATTACAGCTGAAGAGTATCTTCGTACCATAT 567  
Db 171 SerLeuMetLeuAenAlaCysTyrAspValIleSerGluGlyIleProPheSerProTyr 190  
Qy 568 CGATGTGAGTATAGGAGTCTTATCACTATTTTAAAGACCTCAATCTAAATTT 627  
Db 191 IleCysAlaGlyValGlyThrAspLeuIleSerMetPheGluAlaIleAenProLysIle 210  
Qy 628 CTTTACCAGAAATAAGTATAGTTACCTATCACACCAAGAGTCTTCGTACCATAT 687  
Db 211 SerTyrGlnGlyLysLeuGlySerTyrSerIleSerProGluAlaSerValPheVal 230  
Qy 688 GGTGATACTACCATGGCTTCTCTCAACCAACATCTGCTTCA-----GTAACCTTT 747  
Db 251 -----SerThrSerThrLeuThrGlyAenHisPheThrIleValThrLeu 265  
Qy 796 GACGTTGATTAATTTGGCGGAATTTGGAATGAGTTTCACTTC 840  
Db 266 SerValCysHisPheGlyValGluLeuGlyGlyArgPheAenPhe 280

## RESULT 16

US-08-953-326-17  
; Sequence 17, Application US/08953326  
; Patent No. 6251872  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Ganta, Roman R.  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Nyika, Aceme  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Mahan, Suman M.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of  
; TITLE OF INVENTION: Animals and Humans  
; FILE REFERENCE: UF-167C1

; CURRENT APPLICATION NUMBER: US/08/953,326  
; CURRENT FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/953,326  
; EARLIER FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/733,230  
; EARLIER FILING DATE: 1996-10-17  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 17  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
US-08-953-326-17

Alignment Scores:  
Pred. No.: 4e-61 Length: 280  
Score: 603.00 Matches: 133  
Percent Similarity: 63.39% Conservative: 54  
Best Local Similarity: 45.08% Mismatches: 78  
Query Match: 40.72% Indels: 30  
DB: 4 Gaps: 10

US-10-062-624-41 (1-840) x US-08-953-326-17 (1-280)

Qy 1 ATGAATTATAGAAAAATTTCTAGTAAGAGCGGTTAATCTCATTAAATGCTCAATCTTACCA 60  
Db 1 MetAenCysLysLysPhePheIleThrThrLeuValSerLeuMetSerPheLeuPro 20  
Qy 61 TATCAGTCTTTTGCAGATCTCTAGGTTCAAGAACTAATGATAACAAAGAAGGC---TTC 117  
Db 21 GlyIleSerPheSerAlaVal-----GlnAenAenValGlyGlyAenPhe 37  
Qy 118 TACATTAGTCAAGTACAATCAAGTATATCAACATTTTAGAAAAATTTCTCTGCTGAAGAA 177  
Db 38 TyrIleSerGlyLysTyrValProSerValSerHisPheGlyValPheSerAlaLysGln 57  
Qy 178 ACTCCTATTAAATGGAACAAATTTCTCACTAAAGTTTCGGAGTAAAGAAAGATGGT 237  
Db 58 -----GluArgAsnThrThrIleGlyValPheGlyLeuLysGlnAenTrp 72  
Qy 238 GAT-----ATAACAAAAAGAC---GATTTTACAAGAGTAGCTCCAGCGCATTTGAT 285  
Db 73 AspGlySerThrIleSerLysAenSerProGluAenThrPheAenValProAenTyrSer 92  
Qy 286 TTT-----CAAAATAACTTAATCAGGATTTTTCAGGAAGTATTTGTTACTCTATGGAC 339  
Db 93 PheLysTyrGluAenAenProPheLeuGlyPheAlaGlyAlaValGlyTyrLeuMetAen 112  
Qy 340 GGACCAAGATAGACTTGAAGTCTCATATCAACAATTTTAAAGACCTCAATCTAAATTT 399  
Db 113 GlyProArgIleGluLeuGluMetSerTyrGluThrPheAenValLysAenGlnGlyAen 132  
Qy 400 AATGATCTGATTAATGAGTCTCTCAACCAATTTTAAAGACCTCAATCTAAATTT 447  
Db 133 AsnTyrLysAenAenAenAenAenAenAenAenAenAenAenAenAenAenAenAen 150  
Qy 448 AAGATGCAATGGAAGATCAGCAATATAGTACTTAAATATGAGCGGATTAATTTATG 507  
Db 151 LysLeuSerAenAlaGlyAspLysPheValPheLeuLysAenGluGlyLeuAenAspIle 170  
Qy 508 TCATTGATGGTTAATCTTGTATGACATTACAGCTGAAGAGTATCTTCGTACCATAT 567  
Db 171 SerLeuMetLeuAenAlaCysTyrAspValIleSerGluGlyIleProPheSerProTyr 190  
Qy 568 GCATGTGAGTATAGGAGTCTTATCACTATTTTAAAGACCTCAATCTAAATTT 627  
Db 191 IleCysAlaGlyValGlyThrAspLeuIleSerMetPheGluAlaIleAenProLysIle 210  
Qy 628 GCTTACCAGAAATAAGTATAGTTACCTATCACACCAAGAGTCTTCGTACCATAT 687  
Db 211 SerTyrGlnGlyLysLeuGlySerTyrSerIleSerProGluAlaSerValPheVal 230  
Qy 688 GGTGATACTACCATGGCTTCTCTCAACCAACATCTGCTTCA-----GTAACCTCT 747

Db 231 GlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProAlaMetIlePro 250  
QY 748 GTAGTATTAAATGATGCTCTCTCAACACCATCTGCTTCA-----GTAACTCTT 795  
Db 251 -----SerThrSerThrLeuThrGlyAsnHisPheThrIleValThrLeu 265  
QY 796 GACGTTGGATACTTTGGCGGAGAAATTTGGAATGAGGTTCCACCTTC 840  
Db 266 SerValCysHisPheGlyValGluLeuGlyGlyArgPheAsnPhe 280  
RESULT 17  
US-09-660-587-6  
; Sequence 6, Application US/09660587  
; Patent No. 6392023  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: US/09/660,587  
; CURRENT FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 6  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of p28-6 protein  
US-09-660-587-6

Alignment Scores:  
Pred. No.: 6.8e-61 Length: 280  
Score: 601.00 Matches: 129  
Percent Similarity: 61.59% Conservative: 49  
Best Local Similarity: 44.64% Mismatches: 93  
Query Match: 40.58% Indels: 18  
Gaps: 4  
US-10-062-624-41 (1-840) x US-09-660-587-6 (1-280)  
QY 1 ATGAATTATAAGAAATTTCTAGTAAGACGGTAAATCTTAATCTCAATCTTACCA 60  
Db 1 MetAsnCysLysLysIleLeuThrThrAlaLeuMetSerLeuMetTyrrAlaPro 20  
QY 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAAGAACTAATGATATAACAAGAGCTTCTAC 120  
Db 21 SerIleSerPheSerAspThrIle-----GlnAspAsnThrGlySerPheTyrr 37  
QY 121 ATTAGTCAAGACTACAACTCAAGTATATCACACTTTAGAAAATTTCTCTGCTGAAGAACT 180  
Db 38 IleSerGlyLysTyrrValProSerValSerHisPheGlyValPheSerAlaLysGlu--- 56  
QY 181 CCTATTATTAAGAACAAATTTCTCTCACTAAAGATTTTCGGACTAAAGAAAGAT----- 234  
Db 57 -----GluArgAsnSerThrValGlyValPheGlyLeuLysHisAspTrpAsn 72  
QY 235 ---GGTGATATACAAA-----AAAGACGATTTTACAAGAGTAGCTCCAGGCATT 282  
Db 73 GlyGlyThrIleSerAsnSerSerProGluAsnIlePheThrValGlnAsnTyrrSerPhe 92  
QY 283 GATTTTTCAAAATAACTTAATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGACGCGA 342  
Db 93 LysTyrrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrrSerMetGlyGly 112  
QY 343 CCAAGATAGAACTTGAAGCTGCATATCAACAAATTTAATCCAAAACACCGATACAAAT 402  
Db 113 ProArgIleGluLeuGluValLeuTyrrGluThrPheAspValLysAsnGlnAsnAsn 132

QY 403 GATACTGATAATGGTGAATACTAT-----AAACATTTTGCATTATCTCGTAAAGAT 453  
Db 133 TyrLysAsnGlyAlaHisArgTyrrCysAlaLeuSerHisHisSerSerAlaThrSerMet 152  
QY 454 GCAATGGAAGATCAGCAATATGTAGTACTTAAAAATGACGGCATAAATTTTATGTCAATTG 513  
Db 153 SerSerAlaSerAsnLysPheValPheLeuLysAsnGluGlyLeuIleAspLeuSerPhe 172  
QY 514 ATGGTTAATACTTGTCTATGACATTACAGCTGAAGAGATATCTTTCGTACCATATGTCATGT 573  
Db 173 MetIleAsnAlaCysTyrrAspIleIleIleGluGlyMetProPheSerProTyrrIleCys 192  
QY 574 GCAGGTATAGCAGCAGATCTTATCACTATTATTAAGACCTCAATCTCAAAATTTCTCTTAC 633  
Db 193 AlaGlyValGlyThrAspValValSerMetPheGluAlaIleAsnProLysIleSerTyrr 212  
QY 634 CAAGGAAAAATAGGTATTAGTTACCTATCATCACCAAGAGTCTCTGCAATTTATTTGGTGA 693  
Db 213 GlnGlyLysLeuGlyLeuGlyTyrrSerIleSerSerGluAlaSerValPheIleGlyGly 232  
QY 694 TACTACATGCGGTTTATGTAATAATTTGAGAAGATACCTGTAAATAACTCTCTGTAGTA 753  
Db 233 HisPheHisArgValIleGlyAsnGluPheArgAspIleProAlaMetValProSerGly 252  
QY 754 TTAATGATGCTCTCTCAACACCATCTGCTTCAAGTAACTCTTGAGTTGGATACTTTGGC 813  
Db 253 SerAsn---LeuProGluAsnGlnPheAlaIleValThrLeuAsnValCysHisPheGly 271  
QY 814 GGAGAAATTTGGAATGAGGTTCCACCTTC 840  
Db 272 IleGluLeuGlyGlyArgPheAsnPhe 280  
RESULT 18  
US-09-261-358A-6  
; Sequence 6, Application US/09261358A  
; Patent No. 6403780  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP  
; CURRENT APPLICATION NUMBER: US/09/261,358A  
; CURRENT FILING DATE: 1999-03-03  
; PRIOR APPLICATION NUMBER: 09/201,458  
; PRIOR FILING DATE: 1998-11-30  
; NUMBER OF SEQ ID NOS: 33  
; SEQ ID NO 6  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of Eca28SA3 protein  
US-09-261-358A-6  
Alignment Scores:  
Pred. No.: 6.8e-61 Length: 280  
Score: 601.00 Matches: 129  
Percent Similarity: 61.59% Conservative: 49  
Best Local Similarity: 44.64% Mismatches: 93  
Query Match: 40.58% Indels: 18  
Gaps: 4  
US-10-062-624-41 (1-840) x US-09-261-358A-6 (1-280)  
QY 1 ATGAATTATAAGAAATTTCTAGTAAGACGGTAAATCTCAATCTTACCACTTACCA 60  
Db 1 MetAsnCysLysLysIleLeuThrThrAlaLeuMetSerLeuMetTyrrAlaPro 20  
QY 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAAGAACTAATGATATAACAAGAGCTTCTAC 120  
Db 21 SerIleSerPheSerAspThrIle-----GlnAspAsnThrGlySerPheTyrr 37



```
QY 121 ATTAGTCAAGTCAATCCAGTATATCAACATTTAGAAAATCTCTGCTGAGAAACT 180
Db 115 SerGlyLysTyrValProSerValSerHisPheGlyValPheSerAlaLysGlu--- 56
QY 181 CCTATTATGAACAAATCTCTCACTAAAAGTTTCGAGCTAAAGAAAGAT----- 234
Db 57 -----GluArgAsnSerThrValGlyValPheGlyLeuLysHisAspTrpAsn 72
QY 235 ---GGTATATAACAAA-----AAAGACGATTTTACAAGAGTAGTCCAGGCATT 282
Db 73 GlyGlyThrIleSerAsnSerSerProGluAsnIlePheThrValGluAsnTyrSerPhe 92
QY 283 GATTTCCTAAATACTTAATACAGATTTTCAGGAAGTATGTTACTCTATGACGGA 342
Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGlyGly 112
QY 343 CCAGATAGAGACTTGAGCTGCATATCAACAATTTAATCCAAAACACCGATAACAT 402
Db 113 ProArgIleGluLeuGluValLeuTyrGluThrPheAspValLysAsnGlnAsnAsn 132
QY 403 GATACCTGATAATGGTGAATACTAT-----AAACATTTTGCATTATCTCGTAAAGAT 453
Db 133 TyrLysAsnGlyAlaHisArgTyrCysAlaLeuSerHisHisSerSerAlaThrSerMet 152
QY 454 GCAATGGAAGATCAGCAATATGTAGTACTTAAATGACGCGATAACTTTTATGTCATTG 513
Db 153 SerSerAlaSerAsnLysPheValPheLeuLysAsnGluGlyLeuIleAspLeuSerPhe 172
QY 514 ATGGTTAATCTCTGATGATACATCAGCTCAAGAGAGTATCTTTCGTACCATATGTCATG 573
Db 173 MetIleAsnAlaCysTyrAspIleIleGluGlyMetProPheSerProTyrIleCys 192
QY 574 GCAGTATAGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAAATTTGCTTAC 633
Db 193 AlaGlyValGlyThrAspValValSerMetPheGluAlaIleAsnProLysIleSerTyr 212
QY 634 CAAGAAAATAGGTATGTTAGTACCCTATCACACGAGAGTCTCTGCAATTTATTGGTGA 693
Db 213 GlnGlyLysLeuGlyLeuGlyTyrSerIleSerSerGluAlaSerValPheIleGlyGly 232
QY 694 TACTACCATCGGTTATGGTAATAATTTTCAAGAGATACCTCTGAATACTCTCTGTAGTA 753
Db 233 HisPheHisArgValIleGlyAsnGluPheArgAspIleProAlaMetValProSerGly 252
QY 754 TTAATGATGCTCTCAACACACATCTGCTTCACTAGTAACTCTTGGAGTGGATCTTTGGC 813
Db 253 SerAsn---LeuProGluAsnGlnPheAlaIleValThrLeuAsnValCysHisPheGly 271
QY 814 GGAGAAATGGATGAGGTTCCACCTTC 840
Db 272 IleGluLeuGlyGlyArgPheAsnPhe 280
```

## RESULT 19

```
US-09-660-587-2
; Sequence 2, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 2
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
```

```
FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis p28-7 protein
US-09-660-587-2
```

## Alignment Scores:

```
Pred. No.: 1,05e-58 Length: 278
Score: 582.00 Matches: 131
Percent Similarity: 59.93% Conservative: 41
Best Local Similarity: 45.64% Mismatches: 99
Query Match: 39.30% Indels: 16
DB: 4 Gaps: 4
```

US-10-062-624-41 (1-840) x US-09-660-587-2 (1-278)

```
QY 1 ATCAATTTATAAGAAAATTTAGTAAAGAGCGGTTAATCTCATTAATGTCAATCTTACCA 60
Db 1 MetAsnCysLysLysIleLeuIleThrThrAlaLeuIleSerLeuMetTyrSerIlePro 20
QY 61 TATCAGCTCTTTTGCAGATCCTGTAGTTCAGAACTAATGATATACAAAGAGGC----- 114
Db 21 SerIleSerPheSerAspThrIle-----GlnAspGlyAsnMetGlyGlyAsn 36
QY 115 TTCTACATTTAGTCAAGTACAAATCCAGTATATACACATTTTAGAAAAATCTCTGCTGAA 174
Db 37 PheTyrIleSerGlyLysTyrValProSerValSerHisPheGlySerPheAlaLys 56
QY 175 GAAACTCCTATTAAATGAACAAATTTCTCTCACTAAAAAAGTTTTCGGACTAAAGAAAGAT 234
Db 57 Glu-----GluSerLysSerThrValGlyValPheGlyLeuLysHisAsp 71
QY 235 GGTGAT-----ATAACAAAAAAGACGATTTTACAGAGTAGTCTCCAGGC 279
Db 72 TrpAspGlySerProIleLeuLysAsnLysHisAlaAspPheThrValProAsnTyrSer 91
QY 280 ATTGATTTTCAAAATAACTTAATACAGATTTTCAGGAAGTATGTTACTCTATGAC 339
Db 92 PheArgTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGly 111
QY 340 GGACCAAGATAGAACTTGAAGCTGCATATCAACAAATTTAATCCAAAAACACCGATAAC 399
Db 112 GlyProArgIleGluPheGluIleSerTyrGluAlaPheAspValLysSerProAsnIle 131
QY 400 AATGATATGATTAATGGTGAATACTATAAATTTTGCATTTCTCTGTAAGATGCAATG 459
Db 132 AsnTyrGlnAsnAspAlaHisArgTyrCysAlaLeuSerHisHisThrSerAlaAlaMet 151
QY 460 GAAGATCAGCAATATGATGATCTTAAATGACGGCATAACTTTTATGCTCATTCATGTT 519
Db 152 GluAlaAspLysPheValPheLeuLysAsnGluGlyLeuIleAspIleSerLeuAlaIle 171
QY 520 AATACTTGTCTATGACATTCAGCTGAAGAGTATCTTTCTGATCCATATGATGTCAGGT 579
Db 172 AsnAlaCysTyrAspIleIleAsnAspLysValProValSerProTyrIleCysAlaGly 191
QY 580 ATAGGACGAGATCTTATCACTATTTTAAAGACCTCAATCTAAAATTTGCTTACCAAGGA 639
Db 192 IleGlyThrAspLeuIleSerMetPheGluAlaThrSerProLysIleSerTyrGlnGly 211
QY 640 AAAATAGGTATTAGTTACCTCTATCACCCTATCACCAGAGAGTCTCTGCATTTATGTTG 699
Db 212 LysLeuGlyIleSerTyrSerIleAsnProGluThrSerValPheIleGlyGlyHisPhe 231
QY 700 CATGGCTTATGTTAATAATTTGAGAAGATACCTGTAATTAACCTCTGTAGTATTAAAT 759
Db 232 HisArgIleIleGlyAsnGluPheArgAspIleProAlaIleValProSerAsnSerThr 251
QY 760 GATGCTCTCAAAACCATCATCTGCTTCACTTAACCTCTTGAGCTTTGGATATCTTTGG 819
Db 252 ThrIleSerGlyProGlnPheAlaThrValThrLeuAsnValCysHisPheGlyLeuGlu 271
QY 820 ATTGGAATGAGGTTCCACCTTC 840
Db 272 LeuGlyGlyArgPheAsnPhe 278
```



72	TtpAspGlySerProIleuLeuLysAsnLysHisAlaAspPheThrValProAsnTySer	91
280	ATTGATTTTCAAATAACTTAATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGGAC	339
92	PheArgTyrgluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTySerMetGly	111
340	GGACCAAGAAATGAACCTGAAGCTCATATCAACAAATTTATCCAAAAACACCGATAAC	399
112	GlyProArgIleGluPheGluLeuSerTyrgluAlaPheAspValLysSerProAsnIle	131
400	AATGATACCTGATAATGGTGAATACATCTATAAACCATTTTGCATTATCTCGTAAAGATGCAATG	459
132	AsnTyrgluAsnAspAlaHisArgTyrglyAlaLeuSerHisHisThrSerAlaIleMet	151
460	GAAGATCAGCAATATGTAGTACTTAAAAATGACGGCATAACTTTTATGTCAATTGATGGTT	519
152	gluAlaAspLysPheValPheLeuLysAsnGluGlyLeuIleAspIleSerLeuAlaIle	171
520	AATACTTGTCTATGACATTACAGCTCAAGGAGTACTTTCGTACCATATGCATGTGCGAGTT	579
172	AsnAlaCysTyrglyAspIleIleAsnAspLysValProValSerProTyrlleCysAlaGly	191
580	ATAGGAGCAGATCTTATCACTACTTTTTAAAGACCTCAATCTTAAATTTTGGTTTACCAAGGA	639
192	IleGlyThrAspLeuIleSerMetPheGluAlaThrSerProLysIleSerTyrgluGly	211
640	AAATAGGTATTAGTTACCTATCACACGAGAAGTCTCTGCATTATTGGTGGATACTAC	699
212	LysLeuGlyIleSerTyrglySerIleAsnProGluThrSerValPheIleGlyGlyHisPhe	231
700	CATGGCGTTATTGGTAAATAATTTGAGAAGATACCTGTATAACTCTGTAGTATTAAAT	759
232	HisArgIleIleGlyAsnGluPheArgAspIleProAlaIleValProSerAsnSerThr	251
760	GATGCTCCTCAAAACACATCTGCTTCAGTAACTCTTGACGTTGGATATTTGGCGGAGAA	819
252	ThrIleSerGlyProGluPheAlaThrValThrLeuAsnValCysHisPheGlyLeuGlu	271
820	ATTGGAATGAGGTTCACTTC	840
272	LeuGlyGlyArgPheAsnPhe	278

RESULT 22  
US-08-953-326-16'  
; Sequence 16, Application US/089533326  
; Patent No. 6251872  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Ganta, Roman R.  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Nyika, Aceme  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Mahan, Suman M.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of  
; TITLE OF INVENTION: Animals and Humans  
; FILE REFERENCE: UF-167C1  
; CURRENT APPLICATION NUMBER: US/08/953,326  
; CURRENT FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/953,326  
; EARLIER FILING DATE: 1997-10-17  
; EARLIER APPLICATION NUMBER: 08/733,230  
; EARLIER FILING DATE: 1996-10-17  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 16  
; LENGTH: 278  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
US-08-953-326-16  
Alignment Scores: 3.95e-58  
Pred. No.: 278

Score:	577.00	Matches:	126
Percent Similarity:	60.55%	Conservative:	49
Best Local Similarity:	43.60%	Mismatches:	94
Query Match:	38.96%	Indels:	20
DB:	4	Gaps:	7

US-10-062-624-41 (1-840) x US-08-953-326-16 (1-278)

QY	1	ATGAATTATAGAAAAATCTCTAGTAAAGACGCGTTAAATCTCATTAATGTCAATCTTTACCA	60
		:::	
Db	1	MetAenCybLysLysPhePheIleThrAlaLeuValSerLeuMetSerPheLeuPro	20
QY	61	TATCAGTCTTTTGGCAGATCTCTGTAGTTCAAGAACTAATGATAACAAGAAAGC---	117
Db	21	GlyIleSerPheSerAspProVal-----GlnGlyAspAenIleSerGlyAenPhe	37
QY	118	TACATTAGTCAAAAGTCAATCCCAAGTATATACACACTTTAGAAAAATCTCTCTCTGAAGA	177
Db	38	TyrValSerGlyLysTyrMetProSerAlaSerHisPheGlyMetPheSerAlaLysGlu	57
QY	178	ACTCCTATTATGGAACAAATCTCTCTACATAAAAAAGTTTTCCGGACTAAAGAAAGAT---	234
Db	58	-----GluLysAenProThrValAlaLeuTyrGlyLeuLysGlnAspTyr	72
QY	235	-----GGTGATATAACAAAAAGACGATTTTCAAGAGTAGTACGCCGCAAT	282
Db	73	GlucGlyIleSerSerSerHisAenAenHisPheAenAenLysGlyTyrSerPhe	92
QY	283	GATTTTCAAAATACCTTAATATCAGGATTTTCAGGAAGTATTTGGTTACTCTATGACCGGA	342
Db	93	LysTyrGluAenAenProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGlyGly	112
QY	343	CCAAAGATGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAAACACCGATAACAAT	402
Db	113	ProArgValGluPheGluValSerTyrGluThrPheAspValLysAenGlnGlyAenAen	132
QY	403	GATACTGATATGTTGATGATACTATAACAATTTTGCATTATCTCGTAAAGAT-----	453
Db	133	TyrLysAenAenAla-----HisArgTyrCysAlaLeuGlyGlnGlnAenSerGly	150
QY	454	GCAATGAAGATCAGCAATATGTAGTACTTTAAAAATGACGCATAACTTTTATGTCATTG	513
Db	151	IleProLysThrSerLysTyrValLeuLysSerGluGlyLeuLeuAspIleSerPhe	170
QY	514	ATGGTTAATACTTCTGATGACATTACAGCTCAAGGAGTATCTTTCGTACCATCATCATGT	573
Db	171	MetLeuAenAlaCysTyrAspIleIleAenGluSerIleProLeuSerProTyrIleCys	190
QY	574	GCAGGTATAGCAGCAGATCTTATCACATATTTTAAAGACCTCAATCTAAAATTTGCTTAC	633
Db	191	AlaGlyValGlyThrAspLeuIleSerMetPheGluAlaThrAenProLysIleSerTyr	210
QY	634	CAAGGAAAAATAGTATTAGTTACCTTATCACACAGAGTCTCTGCAATTTTATTCGGTGA	693
Db	211	GlnGlyLysLeuGlyLeuSerTyrSerIleAenProGluAlaSerValPheIleGlyGly	230
QY	694	TACTPACCATGCGTTATTGGTAAATAATTTGAGAAGATACCTGTAATAACTCCTCTAGTA	753
Db	231	HisPheHisLysValIleGlyAenGluPheArgAspIleProThrLeuLysAlaPheVal	250
QY	754	TTAAATGATGCTCCTCAAAACACATCTGCTTTCAGTAACTCTTGACGTTGGATCTTTGGC	813
Db	251	ThrSerSerAlaThrProAspLeu--AlaIleValThrLeuSerValCysHisPheGly	269
QY	814	GGAGAAATGGAATGAGGTTCCACCTTC	840
Db	270	IleGluLeuGlyGlyAArgPheAenPhe	278

RESULT 23

US-09-660-587-13

Sequence 13, Application US/09660587

Patent No. 639203

GENERAL INFORMATION:

```
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: US/09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 13
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1E
US-09-660-587-13

Alignment Scores:
Pred. No.: 3,95e-58 Length: 278
Score: 577.00 Matches: 126
Percent Similarity: 60.55% Conservative: 49
Best Local Similarity: 43.60% Mismatches: 94
Query Match: 38.96% Indels: 20
DB: 4 Gaps: 7

US-10-062-624-41 (1-840) x US-09-660-587-13 (1-278)
QY 1 ATGAATTATAAGAAATCTAGTAAGAGCGGTTAATCTCATTATGTCATCTTACCA 60
Db 1 MetAsnCysLysLysPhePheleThrAlaLeuValSerLeuMetSerPheLeuPro 20
QY 61 TATCAGTCTTTTGCAGATCCTCTAGGTTCAAGAACTAATGATACAAAGAGGC---TTC 117
Db 21 GlylleSerPheSerAspProVal-----GlnGlyAspAsnIleSerGlyAsnPhe 37
QY 118 TACATTAGTCAAGATACAAATCCAAAGTATATCACATTTTAGAAAATCTCTGCTGAAGA 177
Db 38 TyrValSerGlyLysTyrMetProSerAlaSerHisPheGlyMetPheSerAlaLysGlu 57
QY 178 ACTCCTATTATGGAACAAATCTCTCACTAAAGAAATTTTCGACTAAAGAAAGAT--- 234
Db 58 -----GlnLysAsnProThrValAlaLeuTyrGlyLeuLysGlnAspTrp 72
QY 235 -----GGTGATATAACAAAAAGAGTATGATGATGATGATGATGATGATGATGAT 282
Db 73 GluGlylleSerSerSerHisAsnAspAsnHisPheAsnAsnLysGlyTyrSerPhe 92
QY 283 GATTTTCAAAATAACTTAATATCAGGATTTTTCAGGAAGTATTGGTTACTTATGACGGA 342
Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGlyGly 112
QY 343 CCAAGATAGAACTTGAAGCTGATATCAACAAATTTAATCCAAAAACACCGATACCAAT 402
Db 113 ProArgValGluPheGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn 132
QY 403 GATCTGTAATGGTGAATACATAACAACTTTTGCATTATCTCGTAAGAT----- 453
Db 133 TyrLysAsnAspAla-----HisArgTyrCysAlaLeuGlyGlnGlnAspAsnSerGly 150
QY 454 GCAATGGAAGATCAGCAATATAGTACTTAAATGAGCGGATACATTTTATGTCATTG 513
Db 151 IleProLysThrSerLysTyrValLeuLeuLysSerGluGlyLeuLeuAspIleSerPhe 170
QY 514 ATGGTTAATACCTGTATGATACATACAGCTGAAGGAGTATCTTTCGATACATGATGAT 573
Db 171 MetLeuAsnAlaCysTyrAspIleleAsnGluSerIleProLeuSerProTyrIleCys 190
QY 574 GAGGTATAGGACGATCTTATCTACTATTTTTAAAGACCTCAATCTAAATTTGCTTAC 633
Db 191 AlaGlyValGlyThrAspLeuIleSerMetPheGluAlaThrAsnProLysIleSerTyr 210
```

```
QY 634 CAAGGAAAAATAGGTATTAGTTACCTATCACACGAGAGTCTCTGCATTTATTGGTGA 693
Db 211 GlnGlyLysLeuGlyLeuSerTyrSerIleAsnProGluAlaSerValPheIleGlyGly 230
QY 694 TACTACCATGCGGTTATTGGTAATAATTTGAGAGATACCTGTAAATAACTCTCTGTAGTA 753
Db 231 HisPheHisLysValIleGlyAsnGluPheArgAspIleProThrLeuLysAlaPheVal 250
QY 754 TTAATGATGCTCTCTCAACACCATCTGCTTCAGTAACTCTTGAGCTTGGATACTTTGGC 813
Db 251 ThrSerSerAlaThrProAspLeu---AlaIleValThrLeuSerValCysHisPheGly 269
QY 814 GGAGAAATTGGAATGAGGTTCCACCTTC 840
Db 270 IleGluLeuGlyGlyArgPheAsnPhe 278

RESULT 24
US-09-261-358A-13
; Sequence 13, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 13
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1E
US-09-261-358A-13
```

```
Alignment Scores:
Pred. No.: 3,95e-58 Length: 278
Score: 577.00 Matches: 126
Percent Similarity: 60.55% Conservative: 49
Best Local Similarity: 43.60% Mismatches: 94
Query Match: 38.96% Indels: 20
DB: 4 Gaps: 7

US-10-062-624-41 (1-840) x US-09-261-358A-13 (1-278)
QY 1 ATGAATTATAAGAAATCTAGTAAGAGCGGTTAATCTCATTATGTCATCTTACCA 60
Db 1 MetAsnCysLysLysPhePheleThrAlaLeuValSerLeuMetSerPheLeuPro 20
QY 61 TATCAGTCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGATACAAAGAGGC---TTC 117
Db 21 GlylleSerPheSerAspProVal-----GlnGlyAspAsnIleSerGlyAsnPhe 37
QY 118 TACATTAGTCAAGATACAAATCCAAAGTATATCACATTTTAGAAAATCTCTGCTGAAGA 177
Db 38 TyrValSerGlyLysTyrMetProSerAlaSerHisPheGlyMetPheSerAlaLysGlu 57
QY 178 ACTCCTATTATGGAACAAATCTCTCACTAAAGAAATTTTCGACTAAAGAAAGAT--- 234
Db 58 -----GlnLysAsnProThrValAlaLeuTyrGlyLeuLysGlnAspTrp 72
QY 235 -----GGTGATATAACAAAAAGAGTATGATGATGATGATGATGATGATGATGAT 282
Db 73 GluGlylleSerSerSerHisAsnAspAsnHisPheAsnAsnLysGlyTyrSerPhe 92
QY 283 GATTTTCAAAATAACTTAATATCAGGATTTTTCAGGAAGTATTGGTTACTTATGACGGA 342
Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGlyGly 112
```

QY	61	TATCAGCTCTTTTCGAGATCTCTGTAGGTTCAAGAACTAATGATACAAAGAAGGC----	TTC	111
Db	21	GlyIleSerPheSerAepProVal-----GlnGlyAepAenIleSerGlyAenPhe	37	
QY	118	TACATTAGTCGAAAGTAGTCAAACTCCAAGTATATCACACTTTAGAAAAATCTCTCGCTGAAGAA	177	
Db	38	TyrValSerGlyIySyrMetProSerAlaSerHisPheGlyMetPheSerAlaIySgLu	57	
QY	178	ACTCCTATTAAATGGGAACAAATTCCTCCTCAATAAAAAAGTTTTCGGAGTAAAGAAAGAT---	234	
Db	58	-----GluLysAenProThrValAlaLeuTyGlyLeuLysGlnAAspTrp	72	
QY	235	-----GGTGATATAACAAAAAAGACGATTTTTCAAGAGTAGTCCAGGCATT	282	
Db	73	GluGlyIleSerSerHisAAsnAAsnHisPheAenAenLysGlyTyrSerPhe	92	
QY	283	GATTTTCAAAATAACTTTAAATATCAGGATTTTTCAGGAAAGTATTGGTTACTCTATGACCGGA	342	
Db	93	LysTyrGluAenAenProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGlyGly	112	
QY	343	CGAAGATAGAACTTGAGCTGCATATCAACAATTTTAATCCAAAAAACACCGATTAACAAAT	402	
Db	113	ProArgValGluPheGluValSerTyrGluThrPheAspValLysAenGlnGlyAenAen	132	
QY	403	GATACTGATTAATGGTGAATACTATAAACATTTTTCGATTATCTCGTAAAGAT-----	453	
Db	133	TyrLysAenAAspAla-----HisArgTyrCysAlaLeuGlyGlnGlnAAsnSerGly	150	
QY	454	GCAATGGAAGATCAGCAATATGTAGTACTTTAAAAATGACGCGATCAACTTTTATGTCATTG	513	
Db	151	IleProLysThrSerLysTyrValLeuLeuLysSerGluGlyLeuLeuAAspIleSerPhe	170	
QY	514	ATGGTTAATACTTGCATATGACATTACAGCTGAAGGAGTATCTTTCGTACCATTCGATGT	573	
Db	171	MetLeuAenAlaCysTyrAspIleIleAenGluSerIleProLeuSerProTyrIleCys	190	
QY	574	GCAGGTATAGGAGCAGATCTTATCACATATTTTAAAGACCTCAATCTAAAAATTTGCTTAC	633	
Db	191	AlaGlyValGlyThrAspLeuIleSerMetPheGluAlaThrAenProLysIleSerTyr	210	
QY	634	CAAGGAAAAATAGGTATTAGTTACCTTATCACACCAGAAAGTCTCTGCACTTTATTGGTGGGA	693	
Db	211	GlnGlyLysLeuGlyLeuSerTyrSerIleAenProGluAlaSerValPheIleGlyGly	230	
QY	694	TACTACCATGGCTTATTGGTAAATAATTTTGAGNAGATACCTGTAATAACTCCTGTAGTA	753	
Db	231	HisPheHisLysValIleGlyAenGluPheArgAspIleProThrLeuLysAlaPheVal	250	
QY	754	TTAAATGATCTCTCTCAAACACATCTCGTTTTCAGTAACTCTTTCAGTTGGATCTTTTCGC	813	
Db	251	ThrSerSerAlaThrProAepLeu--AlaIleValThrLeuSerValCysHisPheGly	269	
QY	814	GGAGAAATTGGATAGGTTTCACCTTC	840	
Db	270	IleGluLeuGlyGlyArgPheAenPhe	278	

574	QY	GCAGGTAACAGAGACAGATTAATACATATTTTAAAGACCTTCAATCTAAAAATTTGCTTAC	633
191	Db	AlaGlyValGlyThrAspLeuIleSerMetPheGluAlaThrAsnProLysIleSerTyr	210
634	QY	CAAGGAAAAATAGGTATTAGTTACCTTATCACACACAGAAGTCTCTGCATTTTATTGGTGG	693
211	Db	GlnGlyLysLeuGlyLeuSerTyrSerIleAsnProGluAlaSerValPheIleGlyGly	230
694	QY	TACTACCATGGCGTTATTGGTATAATAATTTTGAGAAGATCACTGTAATAACTCCTGTACTGA	753
231	Db	HisPheHisLysValIleGlyAsnGluPheArgAspIleProThrLeuLysAlaPheVal	250
754	QY	TTAAATGATGCTCCTCMAACACATCTGCTTTCAGTAACTCTTGACGTTGGATACTTTGGC	813
251	Db	ThrSerSerAlaThrProAspLeu---AlaIleValThrLeuSerValCysHisPheGly	269
814	QY	GGAGAAATTGGAATGAGGTTTCACCTTC	840
270	Db	IleGluLeuGlyGlyArgPheAsnPhe	278

191	AlaGlyValGlyThrAspLeuIleSerMetPheGluAlaThrAsnProLysIleSer	Yr	210
634	CRAGGAAATAGGTATTAGCTTACCTCATCACACAGAGTCTCTGCATTTATTGGTCGA	QY	693
211	GinGlyLysLeuGlyLeuSerIleAsnProGluAlaSerValPheIleGly	Db	230
694	TACTACCATGGCGTTATTGGTAATAATTTGAGAAGATACCTGTAATAACTCTCTAGTA	QY	753
231	HisPheHisLysValIleGlyAsnGluPheArgAspIleProThrLeuLysAlaPheVal	Db	250
754	TTAAATGATGCTCTCTCAAACCATCATCTGCTTCAGTAACCTCTTGACGTTCGATTTGGC	QY	813
251	ThrSerSerAlaThrProaspLeu--AlaIleValThrLeuSerValCysHisPheGly	Db	269
814	GGAGAAATTCGAATGAGTTTCACCTTC	QY	840
270	IleGluLeuGlyGlyArgPheAsnPhe	Db	278

```

/ Sequence 4, Application US/08733230
/ Patent No. 6025338
/
/ GENERAL INFORMATION:
/
/ APPLICANT: Barbet, Anthony F.
/ APPLICANT: Ganta, Roman Reddy
/ APPLICANT: McGuire, Travis C.
/ APPLICANT: Burridge, Michael J.
/ APPLICANT: Nyika, Aceme
/ APPLICANT: Kurangirwa, Fred R.
/ APPLICANT: Mahan, Suman M.
/
/ TITLE OF INVENTION: Nucleic Acid Vaccines Against
/ TITLE OF INVENTION: Rickettsial Diseases and Methods of Use
/
/ NUMBER OF SEQUENCES: 6
/
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Saliwanchik & Saliwanchik
/

```

APPLICANT: Rurangirwa, Fred R.  
APPLICANT: Mahan, Suman M.  
TITLE OF INVENTION: Nucleic Acid Vaccines Against  
TITLE OF INVENTION: Rickettsial Diseases and Methods of Use  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saliwanchik & Saliwanchik

```
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/733,230
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF-167
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 280 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-733-230-4

Alignment Scores:
Pred. No.: 7,34e-57 Length: 280
Score: 566.00 Matches: 130
Percent Similarity: 59.04% Conservative: 43
Best Local Similarity: 44.37% Mismatches: 92
Query Match: 38.22% Indels: 28
DB: 3 Gaps: 8

US-10-062-624-41 (1-840) x US-08-733-230-4 (1-280)
QY 1 ATGAATTATAAGAAATTCAGTAAGACGGCGTTAATCTCAATTAATGTCAATCTTACCA 60
DB 1 MetAsnTyrLysSerPheIleThrAlaIleAspIleAsnIleLeuLeuPro 20
QY 61 TATCAGTCTTTGCAGATCTCTAGTTCACAA-----ACTAATGATACAAA 108
DB 21 GlyValSerPheSerAspPro-----ArgGlnValValIleAsnGlyLeuLys 36
QY 109 GAAGGCTTCTACATTAGTCAAGTACAAATCCAAATATATACACTTTAGAAATTTCTCT 168
DB 37 -----PheTyrIleSerGlyLysTyrAspAlaLysAlaSerHisPheGlyValPheSer 54
QY 169 GCTGAAGAAACTCTTATTAAATGGAACAAATCTCTCACTAAAGTTTTCGAGCTAAAG 228
DB 55 AlaLysGlu-----GluArgAsnThrThrValGlyValPheGlyLeuLys 69
QY 229 AAAGATGGTGAT-----ATAACAAAAAGACGATTTTACAGAGTA 270
DB 70 GlnAsnTyrAspGlySerAlaIleSerAsnSerProAsnAspValPheThrValSer 89
QY 271 GCTCCAGGCGATTGTTTCAAAATTAATCAATCAGGATTTTCAGGAAGTATGTTTAC 330
DB 90 AsnTyrSerPheLysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyr 109
QY 331 TCTATGGACGGACCAAGATAGAACTTGAAGTGCATATCAACAAATTAATCCAAAAAC 390
DB 110 SerMetAspGlyProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsn 129
QY 391 ACCGATACAAATGACTACTGATATGCTGAATACTATAACATTTGCCATATCTCGTAA 450
DB 130 GlnGlyAsnAsn-----TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerHisAsn 147
QY 451 GATGCAATGGAA-----GATCAGCAATATGTAGTACTTTAAAAATGACGCGATA 498
DB 111 -----GlnGlyAsnGluAlaHisArgTyrCysAlaLeuSerHisAsn 147

148 SerAlaAlaAspMetSerSerAlaSerAsnAsnPheValPheLeuLysAsnGluGlyLeu 167
499 ACTTTTATGTCATTGATGGTTAATACTTGTATGACATTACAGCTGAAGGAGTATCTTTC 558
168 LeuAspIleSerPheMetLeuAsnAlaCysTyrAspValValGlyGlyLeuProPhe 187
559 GTACCATATGCTGTCAGGATATAGGACGACGATCTTACACTATTTTAAAGACCTCAAT 618
188 SerProTyrIleCysAlaGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsn 207
619 CTAATAATTTGCTTACCAAGGAAAAATAGTATTAGTTTACCCCTATCACACCAAGAGTCTCT 678
208 ProLysIleSerTyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSer 227
679 GCATTATTTGTCGATCTACTACCATGGCGTTATTTGGTAAATAATTTGAGAAGATACCTGTA 738
228 ValPheIleGlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThr 247
739 ATRAACCTCTGTAGTATTAAATGATGCTCCTCAA---ACCACATCTGCTTCAGTAACTCTT 795
248 IleIleProThrGlySerThrLeuAlaGlyLysGlyAsnTyrProAlaIleValIleLeu 267
796 GACGTTGCGATCTTTGCGGAGAAATTCGAATGAGGTTTC 834
268 AspValCysHisPheGlyIleGluMetGlyGlyArgPhe 280

RESULT 27
US-08-953-326-4
; Sequence 4, Application US/08953326
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-4

Alignment Scores:
Pred. No.: 7,34e-57 Length: 280
Score: 566.00 Matches: 130
Percent Similarity: 59.04% Conservative: 43
Best Local Similarity: 44.37% Mismatches: 92
Query Match: 38.22% Indels: 28
DB: 4 Gaps: 8

US-10-062-624-41 (1-840) x US-08-953-326-4 (1-280)
QY 1 ATGAATTATAAGAAATTCAGTAAGACGGCGTTAATCTCAATTAATGTCAATCTTACCA 60
DB 1 MetAsnTyrLysSerPheIleThrAlaIleAspIleAsnIleLeuLeuPro 20
QY 61 TATCAGTCTTTGCAGATCTCTAGTTCACAA-----ACTAATGATACAAA 108
DB 21 GlyValSerPheSerAspPro-----ArgGlnValValIleAsnGlyLeuLys 36
```

```

QY 109 GAAGCTTCTACATTAGTCAAGTACAAATCCAGTATATACACATTCTAGAAAATCTCT 168
Db 37 -----PheTyrIleSerGlyLysAspAlaLysAlaSerHisPheGlyValPheSer 54
QY 169 GCTGAAGAACTCCTATTAAATGGAACAAATCTCTCACTAAAGAGTTTCGGACTAAG 228
Db 55 AlalysGlu-----GluArgAsnThrThrValGlyValPheGlyLeuLys 69
QY 229 AAAGATGGTGAT-----ATAACAAAAGAGAGTATTTTACAGAGTA 270
Db 70 GlnAsnTrpAspGlySerAlaIleSerAsnSerProAsnAspValPheThrValSer 89
QY 271 GCTCCAGCATGATTTTCAAAATTAATATATACAGATTTTCAGGAAGATTGGTTAC 330
Db 90 AsnTyrSerPheLysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyr 109
QY 331 TCTATGACGACCAAGAAATAGACTTCAAGCTGCATATCAACAAATTTAAATCCAAAAC 390
Db 110 SerMetAspGlyProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsn 129
QY 391 ACCGATAACAATGATGATAATGCTGAATGCTATATAACAACTTTTGCATTATCTCGTAA 450
Db 130 GlnGlyAsnAsn-----TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerHisAsn 147
QY 451 GATGCAATGGAA-----GATCAGCAATATGATGATCTTAAATATGACGGCATA 498
Db 148 SerAlaAlaAspMetSerSerAlaSerAsnAsnPheValPheLeuLysAsnGluGlyLeu 167
QY 499 ACTTTATGTCATGCTGTTAAATGCTGATGATGATGATGATGATGATGATGATGATG 558
Db 168 LeuAspIleSerPheMetLeuAsnAlaCysTyrAspValValGlyGluGlyIleProPhe 187
QY 559 GTACCATATGATGTCAGGATATAGGACGATCTTATCACTATTTTAAAGACCTCAAT 618
Db 188 SerProTyrIleCysAlaGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsn 207
QY 619 CTAAATTTGCTTACCAAGAAATAGTATAGTATAGTATAGTATAGTATAGTATAGTAT 678
Db 208 ProLysIleSerTyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSer 227
QY 679 GCATTATTTGCTGATCTACCATGCGGTATTTGCTATTAATTTTGAAGATACCTGTA 738
Db 228 ValPheIleGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThr 247
QY 739 ATAACCTCTGATGATTAATGATGCTCTCTCA---ACCACATCTGCTTCAAGTAACTCT 795
Db 248 IleIleProThrGlySerThrLeuAlaGlyLysGlyAsnTyrProAlaIleValIleLeu 267
QY 796 GACGTGATGATCTTGGCGGAATTTGGAATGAGGTTTC 834
Db 268 AspValCysHisPheGlyIleGluMetGlyGlyArgPhe 280

```

RESULT 28

```

US-09-660-587-15
; Sequence 15, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 15
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
; FEATURE:

```

OTHER INFORMATION: amino acid sequence of C. ruminantium MAP-1  
US-09-660-587-15

Alignment Scores:  
Pred. No.: 9,63e-57 Length: 284  
Score: 565.00 Matches: 129  
Percent Similarity: 56.19% Conservative: 39  
Best Local Similarity: 43.14% Mismatches: 97  
Query Match: 38.15% Indels: 34  
DB: 4 Gaps: 6

US-10-062-624-41 (1-840) x US-09-660-587-15 (1-284)

```

QY 1 ATGAATTATAGAAAAATTTCTAGTAGAAGCGCGTAAATCTCATTAATGTCATCTTACCA 60
Db 1 MetAsnCysLysLysIlePheIleThrSerThrLeuIleSerLeuValSerPheLeuPro 20
QY 61 TATCAGCTCTTTTCAGAT-----CCTGTAGGTTCAAGAACTAAT 99
Db 21 GlyValSerPheSerAspValIleGlnGluAsnAsnProValGlySer----- 37
QY 100 GATAACAAAGAGGCTTCTACATTAGTGCAAGTACAAATCCCAAGTATATACACTTTAGA 159
Db 38 -----ValTyrIleSerAlaLysTyrMetProThrAlaSerHisPheGly 52
QY 160 AAATTCTCTGCTGAAGAACTCTCTATTAAATGGAAACAAATCTCTCACTAAAGAGTTTTC 219
Db 53 LysMetSerIleLysGluAspSerArgAsp-----ThrLysAlaValPhe 67
QY 220 GGACTAAAGAAAGATGCTGATATATAACAAAAAGACGAT-----TTT 261
Db 68 GlyLeuLysLysAspTyrAspGlyValLysThrProSerGlyAsnThrAsnSerIlePhe 87
QY 262 ACAAGAGTAGCTCCAGCAGTATTGATTTTCAAAATTAATATATCAGGATTTTCAGGAAGT 321
Db 88 ThrGluLysAspTyrSerPheLysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAla 107
QY 322 ATTGGTTACTTATGGACGACCAAGATAGAACTTGAAGCTGCATATCAACAATTAAT 381
Db 108 ValGlyTyrSerMetAsnGlyProArgIleGluPheGluValSerTyrGluThrPheAsp 127
QY 382 CCAAAAAACCCGATAACAATGATAGTACTGTAATGCTGAATACTATAACATTTTGCATTA 441
Db 128 ValArgAsnProGlyGlyAsnTyrLysAsnAspAlaHisMetTyrCysAlaLeuAspThr 147
QY 442 TCTCGTAAAGATGCAATGGAAGATCAGCAATATGCTAGTACTTAAATATGACGGCATACT 501
Db 148 AlaSerSerThrAlaGlyAlaThrSerValMetValLysAsnGluAsnLeuThr 167
QY 502 TTTATGTCATTGATGTTAATATCTGCTATGACATTAACAGCTGAAGGAGTATCTTCGTA 561
Db 168 AspIleSerLeuMetLeuAsnAlaCysTyrAspIleMetLeuAspGlyMetProValSer 187
QY 562 CCATATGCTGTCAGGATATAGGACGATCTTATCACTATTTTAAAGACCTCAATCTA 621
Db 188 ProTyrValCysAlaGlyIleGlyThrAspLeuValSerValIleAsnAlaThrAsnPro 207
QY 622 AAATTTGCTTACCAAGAAATAGTATTAGTTACCTTATACACACAGAGTCTCTGCA 681
Db 208 LysLeuSerTyrGlnGlyLysLeuGlyIleSerTyrSerIleAsnProGluAlaSerIle 227
QY 682 TTTATTTGCTGATCTACCATGCGTTATTGGTAAATAAATTTGAGAAGATACCTGTAATA 741
Db 228 PheIleGlyGlyHisPheHisArgValIleGlyAsnGluPheLysAspIle-----Ala 245
QY 742 ACTCCTGTAGTATTAAATGATGCTCTCAACACCATCT-----GCT 783
Db 246 ThrSerLysValPheThrSerSerGlyAsnAlaSerAlaValSerProGlyPheAla 265
QY 784 TCAGTAACTCTTGACGTTGGTACTTTTGGCGGAGAAATTTGGAATGAGGTTTCACCTTC 840
Db 266 SerAlaIleLeuAspValCysHisPheGlyIleGluIleGlyArgPheValPhe 284

```





```

Qy 262 ACAAGAGTAGCTCCAGCAATGATTTTCAAAATAAATCACTTAATATACAGGATTTTCAGGAAGT 321
Db ThrGluLysAspTyrSerPheLysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAla 107
Qy 322 ATTGGTTACTCTATGGACGACCAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 381
Db ValGlyTyrSerMetAsnGlyProArgGluPheGluValSerTyrGluThrPheAsp 127
Qy 382 CCAAAAAACACCGATAACAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 441
Db ValArgAsnProGlyGlyAsnTyrLysAsnAspAlaHisMetTyrCysAlaLeuAspThr 147
Qy 442 TCTCGTAAAGATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 501
Db AlaSerSerThrAlaGlyAlaThrThrSerValMetValLysAsnGluAsnLeuThr 167
Qy 502 TTTATGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 561
Db AspLeuSerMetLeuAsnAlaCysTyrAspLeuMetLeuAspGlyMetProValSer 187
Qy 562 CCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 621
Db ProTyrValCysAlaGlyIleGlyThrAspLeuValSerValIleAsnAlaThrAsnPro 207
Qy 622 AAATTTGCTTACCAAGGAAATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTAT 681
Db LysLeuSerTyrGlnGlyLysLeuGlyLysLeuGlyLysLeuGlyLysLeuGlyLysLeu 227
Qy 682 TTTATGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 741
Db PheLeuGlyLysPheHisArgValIleGlyAsnGluPheLysAspIle-----Ala 245
Qy 742 ACTCCTGTAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 783
Db ThrSerLysValPheThrSerSerGlyAsnAlaSerSerAlaValSerProGlyPheAla 265
Qy 784 TCAGTAATCTTTCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 840
Db SerAlaIleLeuAspValCysHisPheGlyIleGluIleGlyLysGlyLysGlyLysGly 284

RESULT 31
US-09-660-587-11
; Sequence 11, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 11
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1C
US-09-660-587-11

Alignment Scores:
Pred. No.: 1.63e-56 Length: 280
Score: 563.00 Matches: 124
Percent Similarity: 59.73% Conservative: 51
Best Local Similarity: 42.32% Mismatches: 92
Query Match: 38.01% Indels: 26
DB: 4 Gaps: 8

US-10-062-624-41 (1-840) x US-09-660-587-11 (1-280)

```

```

Qy 1 ATGAATTATAAGAAAATTTCTAGTAAGAAGCGCGTAAATCTCATTAATGTCAATCTTACCA 60
Db MetAsnCysLysPhePheIleThrAlaLeuAlaLeuProMetSerPheLeuPro 20
Qy 61 TATCAGTCTTTTTCAGATCTCTAGTGTTCAGAACTAATGATAACAAGAAGGC---TTC 117
Db GlyLeuLeuSerGluProVal-----GlnAspAspSerValSerGlyAsnPhe 37
Qy 118 TACATTAGTCAAAAGTACAATCAAGTATATACACATTTTAGMAAATTTCTCTGCTGAAGAA 177
Db TyrIleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu 57
Qy 178 ACTCCTATTAAATGAACAATTTCTCTCACTAAAAAGTTTTCGACTAAAGAAAGATGGT 237
Db 58 -----GluLysAsnProThrValAlaLeuTyrGlyLeuLysGlnAspTrp 72
Qy 238 GATATAACAACAAAAGAC-----GATTTTACAGAGTAGTCTCCAGGCATT 282
Db 73 AsnGlyValSerAlaSerSerHisAlaAspAlaAspPheAsnAsnLysGlyTyrSerPhe 92
Qy 283 GATTTTCAAAATAACTTAATATCAGGATTTTCAGGAAGTATTTGGTTACTCTATGACGGA 342
Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGlyGly 112
Qy 343 CCAAGAATAGAACTTGAAGCTGCATATCAACAATTTTAAATCCAAAAACACCGATAACAAT 402
Db 113 ProArgIleGluPheGluValSerTyrGluThrPheAspValLysAsnGlnGlyGlyAsn 132
Qy 403 GATCTGATTAATGGTGAATACTATAACATTTTGCATTTCTCGTAAAGATGCAATGAA 462
Db 133 TyrLysAsnAspAla-----HisArgTyrCysAlaLeuAspArgLysAlaSerSerThr 150
Qy 463 GAT-----CAGCAATATGATAGTACTTAAATAATCAGCGCATACATTTTATCTCATTTG 513
Db 151 AsnAlaThrAlaSerHisTyrValLeuLeuLysAsnGluGlyLeuLeuAspIleSerLeu 170
Qy 514 ATGGTTAATATCTGCTATGACATTTACAGCTGAAGAGGATATCTTTTCGTACCATATGATGT 573
Db 171 MetLeuAsnAlaCysTyrAspValValSerGluGlyIleProPheSerProTyrIleCys 190
Qy 574 GCAGGTATAGGAGCATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTAC 633
Db 191 AlaGlyValGlyThrAspLeuIleSerMetPheGluAlaIleAsnProLysIleSerTyr 210
Qy 634 CAAGGAAAAATAGTATTAGTCTACCTATCACACCAAGAGTCTCTGCAATTTATTGTTGGA 693
Db 211 GlnGlyLysLeuGlyLeuSerTyrSerIleAsnProGluAlaSerValPheValGlyGly 230
Qy 694 TACTACCATGGCGTTATTGTTGAATAAATTTGAGAAGTACCTGTAAATA----- 741
Db 231 HisPheHisLysValAlaGlyAsnGluPheArgAspIleSerThrLeuLysAlaPheAla 250
Qy 742 ACTCCTGTAGTATTAATGATGCTCTCTCAACACCATCTCTGCTCAGTAACCTTTGAGTT 801
Db 251 ThrPro-----SerSerAlaAlaThrProAspLeuAlaThrValThrLeuSerVal 267
Qy 802 GGATACTTTGCGGAGAAATTTGGAATGAGGTTACCTTCACCTTC 840
Db 268 CysHisPheGlyValGluLeuGlyGlyArgPheAsnPhe 280

RESULT 32
US-09-261-358A-11
; Sequence 11, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A

```

```
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 11
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1C
US-09-261-358A-11

Alignment Scores:
Pred. No.: 1.63e-56 Length: 280
Score: 563.00 Matches: 124
Percent Similarity: 59.73% Conservative: 51
Best Local Similarity: 42.32% Mismatches: 92
Query Match: 38.01% Indels: 26
DB: 4 Gaps: 8

US-10-062-624-41 (1-840) x US-09-261-358A-11 (1-280)
Qy 1 ATGAATTATAAGAAATTCCTAGTAAGAGCGGTTAAATCTCATTAAATGTCATCTTACCA 60
Db 1 MetAsnCysLysPhePheIleThrAlaLeuAlaLeuProMetSerPheLeuPro 20
Qy 61 TATCAGCTCTTTGCGAGATCCTGTAGGTCGAAGAACTAATGATATACAAAGAAGGC---TTC 117
Db 21 GlyIleLeuLeuSerGluProVal-----GlnAspAspSerValSerGlyAsnPhe 37
Qy 118 TACATTAGTCGAAGTACCAATCCAGTATATACACACTTTAGAAAATTCCTCGCTGAAGAA 177
Db 38 TyrIleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu 57
Qy 178 ACTCCTATTAAATGGAACAAATCTCTCACTAAAAAGTTTTCGGACTAAAGAAAGATGGT 237
Db 58 -----GluLysAsnProThrValAlaLeuTyrGlyLeuLysGlnAspTrp 72
Qy 238 GATATAACAAAAAAGAC-----GATTTTACAGAGTAGCTCCAGGCATT 282
Db 73 AsnGlyValSerAlaSerHisAlaAspPheAsnAsnLysGlyTyrSerPhe 92
Qy 283 GATTTCAAAATACTTAATATCAGAGTATTCAGGAAGTATTTGGTTACTCTCTGACCGGA 342
Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGlyGly 112
Qy 343 CCAAGATAGAACTTTGAAGTCGCATATCAACAATTTTAAATCCAAAAAACCCGATAAACAAT 402
Db 113 ProArgIleGluPheGluValSerTyrGluThrPheAspValLysAsnGlnGlyGlyAsn 132
Qy 403 GATACCTGATATGTTGAATATCTATAAACAATTTTGCATTATCTCGTAAAGATGCAATGGA 462
Db 133 TyrLysAsnAspAla-----HisArgTyrCysAlaLeuAspArgLysAlaSerThr 150
Qy 463 GAT-----CACCAATATCTAGTACTTAAATAATGACGGCATACCTTTATGTCATTG 513
Db 151 AsnAlaThrAlaSerHisTyrValLeuLysAsnGluGlyLeuLeuAspLysSerLeu 170
Qy 514 ATGGTTAATCTGCTATGACATACAGCTGAAGGAGTATCTTCGTACCATATCATGCT 573
Db 171 MetLeuAsnAlaCysTyrAspValSerGluGlyIleProPheSerProTyrIleCys 190
Qy 574 GCAGGTATAGGAGATCTTATCACTATTATTTAAAGACCTCAATCTCAAAATTCCTTAC 633
Db 191 AlaGlyValGlyThrAspLeuIleSerMetPheGluAlaIleAsnProLysIleSerTyr 210
Qy 634 CAAGGAAAAATAGGTATTAGTTACCTCATCACACAGAGTCTCTGCACTTTATTCGTGGA 693
Db 211 GlnGlyLysLeuGlyLeuSerTyrSerIleAsnProGluAlaSerValPheValGlyGly 230
Qy 694 TACTACCATGGCGTTATTGGTAAATAAATTTGAGAAGATACCTGTTAATA-----741
Db 231 HisPheHisLysValAlaGlyAsnGluPheArgAspIleSerThrLeuLysAlaPheAla 250

US-10-062-624-41 (1-840) x US-09-201-458-7 (1-280)
Qy 1 ATGAATTATAAGAAATTCCTAGTAAGAGCGGTTAAATCTCATTAAATGTCATCTTACCA 60
Db 1 MetAsnCysLysPhePheIleThrAlaLeuAlaLeuProMetSerPheLeuPro 20
Qy 61 TATCAGCTCTTTGCGAGATCCTGTAGGTCGAAGAACTAATGATATACAAAGAAGGC---TTC 117
Db 21 GlyIleLeuLeuSerGluProVal-----GlnAspAspSerValSerGlyAsnPhe 37
Qy 118 TACATTAGTCGAAGTACCAATCCAGTATATACACACTTTAGAAAATTCCTCGCTGAAGAA 177
Db 38 TyrIleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu 57
Qy 178 ACTCCTATTAAATGGAACAAATCTCTCACTAAAAAGTTTTCGGACTAAAGAAAGATGGT 237
Db 58 -----GluLysAsnProThrValAlaLeuTyrGlyLeuLysGlnAspTrp 72
Qy 238 GATATAACAAAAAAGAC-----GATTTTACAGAGTAGCTCCAGGCATT 282
Db 73 AsnGlyValSerAlaSerHisAlaAspPheAsnAsnLysGlyTyrSerPhe 92
Qy 283 GATTTCAAAATACTTAATATCAGAGTATTCAGGAAGTATTTGGTTACTCTCTGACCGGA 342
Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGlyGly 112
Qy 343 CCAAGATAGAACTTTGAAGTCGCATATCAACAATTTTAAATCCAAAAAACCCGATAAACAAT 402
Db 113 ProArgIleGluPheGluValSerTyrGluThrPheAspValLysAsnGlnGlyGlyAsn 132
Qy 403 GATACCTGATATGTTGAATATCTATAAACAATTTTGCATTATCTCGTAAAGATGCAATGGA 462
Db 133 TyrLysAsnAspAla-----HisArgTyrCysAlaLeuAspArgLysAlaSerThr 150
Qy 463 GAT-----CACCAATATCTAGTACTTAAATAATGACGGCATACCTTTATGTCATTG 513
Db 151 AsnAlaThrAlaSerHisTyrValLeuLysAsnGluGlyLeuLeuAspLysSerLeu 170
Qy 514 ATGGTTAATCTGCTATGACATACAGCTGAAGGAGTATCTTCGTACCATATCATGCT 573
Db 171 MetLeuAsnAlaCysTyrAspValSerGluGlyIleProPheSerProTyrIleCys 190
Qy 574 GCAGGTATAGGAGATCTTATCACTATTATTTAAAGACCTCAATCTCAAAATTCCTTAC 633
Db 191 AlaGlyValGlyThrAspLeuIleSerMetPheGluAlaIleAsnProLysIleSerTyr 210
Qy 634 CAAGGAAAAATAGGTATTAGTTACCTCATCACACAGAGTCTCTGCACTTTATTCGTGGA 693
Db 211 GlnGlyLysLeuGlyLeuSerTyrSerIleAsnProGluAlaSerValPheValGlyGly 230
Qy 694 TACTACCATGGCGTTATTGGTAAATAAATTTGAGAAGATACCTGTTAATA-----741
Db 231 HisPheHisLysValAlaGlyAsnGluPheArgAspIleSerThrLeuLysAlaPheAla 250
```

Qy 463 GAT-----CAGCAATATGTAGTACTTAAATGACGCGCATAACTTTTATGCTCATG 513  
Db 151 AsnAlaThrAlaSerHisTyrValLeuLeuLysAsnGluGlyLeuLeuAspIleSerLeu 170  
Qy 514 ATGGTTAATCTGCTATGACATTACAGCTGAAGAGTATCTTTTCGTACCATATGCGATGT 573  
Db 171 MetLeuAsnAlaCysTyrAspValSerGluGlyIleProPheSerProTyrIleCys 190  
Qy 574 GCAGGTATGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTTAAATTTGCTTAC 633  
Db 191 AlaGlyValGlyThrAspLeuIleSerMetPheGluAlaIleAsnProLysIleSerTyr 210  
Qy 634 CAAGCAAAATAGTATATGTTACCTTATCACACAGAGTCTCTGCAATTTATGTTGCGA 693  
Db 211 GlnGlyLysLeuGlyLysSerTyrSerIleAsnProGluAlaSerValPheValGlyGly 230  
Qy 694 TACTACCATGCGCTTATTGGTAATAATTTGAGAAGATACCTGTAATA-----741  
Db 231 HisPheHisLysValAlaGlyAsnGluPheArgAspIleSerThrLeuLysAlaPheAla 250  
Qy 742 ACTCCTGTAGTATTAATGATGCTCCTCAACACCATCTGCTTCAGTAACCTCTTTCAGGTT 801  
Db 251 ThrPro-----SerSerAlaAlaThrProAspLeuAlaThrValThrLeuSerVal 267  
Qy 802 GGATACTTTCGCGAGAAATTTGGAATGAGTTTCACCTTC 840  
Db 268 CysHisPheGlyValGluLeuGlyGlyArgPheAsnPhe 280

RESULT 34

US-08-733-230-2  
; Sequence 2, Application US/08733230  
; Patent No. 6025338  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Ganta, Roman Reddy  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Nyika, Aceme  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Mahan, Suman M.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against  
; TITLE OF INVENTION: Rickettsial Diseases and Methods of Use  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: US/08/733,230  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Whitlock, Ted W.  
; REGISTRATION NUMBER: 36,965  
; REFERENCE/DOCKET NUMBER: UP-167  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 352-375-8100  
; TELEFAX: 352-372-5800  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 287 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-733-230-2

Alignment Scores: 7.08e-56 Length: 287  
Pred. No.: 557.50 Matches: 127  
Score: 57.14% Conservative: 41  
Best Local Similarity: 43.20% Mismatches: 105  
Query Match: 37.64% Indels: 21  
DB: 3 Gaps: 6  
US-10-062-624-41 (1-840) x US-08-733-230-2 (1-287)  
Qy 1 ATGATTTATAGAAATTTCTAGTAAGAGCGCGTTAATCTCATTAATGTCATCTTACCA 60  
Db 1 MetAsnCysLysLysIlePheIleThrSerThrLeuIleSerLeuValSerPheLeuPro 20  
Qy 61 TATCAGCTTTTTCAGATCTCTGAGGTTCAGAACTAATGATACACAAAGAGGC---TTC 117  
Db 21 GlyValSerPheSerAspValIle-----GlnGluAspSerAsnProAlaGlySerVal 38  
Qy 118 TACATTAAGTCAAGTACAAATCCAAAGTATATACACTTTAGAAAATTTCTCTGCTGAAGA 177  
Db 39 TyrIleSerAlaLysTyrMetProThrAlaSerHisPheGlyLysMetSerIleLysGlu 58  
Qy 178 ACTCCTATTATGGAACAAATTTCTCTCACTAAAAAGTTTTCGACTAAAGAAAGATGGT 237  
Db 59 AspSerLysAsn-----ThrGlnThrValPheGlyLeuLysLysAspTrp 73  
Qy 238 GATATATAACAAAAAAGACGAT-----TTTACAAGAGTA 270  
Db 74 AspGlyValLysThrProSerAspSerAsnThrAsnSerThrIlePheThrGluLys 93  
Qy 271 GCTCCAGGCATTCATTTTCAATAACTTAATATCAGGATTTTCAGGAAGTATTTGGTTAC 330  
Db 94 AspTyrSerPheArgTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyr 113  
Qy 331 TCTATGACGCGCAAGAAATAGAACTTGAAGCTGCATATCAACATTTTATATCCAAAAAAC 390  
Db 114 SerMetAsnGlyProArgIleGluPheGluValSerTyrGluThrPheAspValLysAsn 133  
Qy 391 ACCGATAACAATGATACATGATAATGTTGGTGAATACATAAACAATTTTGCATTATCTCGTAAA 450  
Db 134 LeuGlyGlyAsnTyrLysAsnAsnAlaHisMetTyrCysAlaLeuAspThrAlaAlaGln 153  
Qy 451 GATGCAATGGAAGAT-----CAGCAATATGTAGTACTTAAATGACGCGCATAACT 501  
Db 154 AsnSerThrAsnGlyAlaGlyLeuThrThrSerValMetValLysAsnGluAsnLeuThr 173  
Qy 502 TTTATGTCATTTGATGTTAATCTGCTATGATACATTTACAGCTGAAGAGTATCTTTCGTA 561  
Db 174 AsnIleSerLeuMetLeuAsnAlaCysTyrAspIleMetLeuAspGlyIleProValSer 193  
Qy 562 CCATATGCTATGTCAGGTATAGGACAGATCTTATCACTATTTTAAAGACCTCAATCTA 621  
Db 194 ProTyrValCysAlaGlyIleGlyThrAspLeuValSerValIleAsnAlaThrAsnPro 213  
Qy 622 AAATTTGCTTACCAAGGAAAAATAGGTATTAGTTACCTATCACACAGAGTCTCTGCA 681  
Db 214 LysLeuSerTyrGlnGlyLysLeuGlyIleSerTyrSerIleAsnSerGluAlaSerIle 233  
Qy 682 TTTATGCTGATACTACCATGGGTTATTGGTAAATAAATTTGAGAGATACCTGTAATA 741  
Db 234 PheIleGlyGlyHisPheHisArgValIleGlyAsnGluPheLysAspIleAlaThrLeu 253  
Qy 742 ACTCCTGTAGTATTAAATGATGCTCCTCAACACCATCT---GCTTCAGTAACCTCTTGAC 798  
Db 254 LysIlePheThrSerLysThrGlyIleSerAsnProGlyPheAlaSerAlaThrLeuAsp 273  
Qy 799 GTTGGATACTTTGCGGAGAAATTTGGAATGAGTTTCACCTTC 840  
Db 274 ValCysHisPheGlyIleGluIleGlyGlyArgPheValPhe 287

RESULT 35

US-08-953-326-2

```
; Sequence 2, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Kurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UP-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 2
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
US-08-953-326-2

Alignment Scores:
Pred. No.: 7,08e-56 Length: 287
Score: 557.50 Matches: 127
Percent Similarity: 57.14% Conservative: 41.
Best Local Similarity: 43.20% Mismatches: 105
Query Match: 37.64% Indels: 21
DB: 4 Gaps: 6

US-10-062-624-41 (1-840) x US-08-953-326-2 (1-287)
QY 1 ATGAATTATAGAAAATCTAGTAGAGCGGTTAAATCTAATTAATCTAATCTTACCA 60
Db 1 MetAsnCysLysLeuValPheThrSerLeuValSerPheLeuPro 20
QY 61 TATCAGTCTTTGAGATCTGTAGGTTCAAGAACTAATGATAACAAAGAGCG---TTC 117
Db 21 GlyValSerPheSerAspValle-----GlnGluAspSerAsnProAlaGlySerVal 38
QY 118 TACATTAGTCAAGTACAATCAAGTATATACACTTTAGAAAATCTCTGCTGAAGAA 177
Db 39 TyrIleSerAlaLysTyrMetProThrAlaSerHisPheGlyLysMetSerIleLysGlu 58
QY 178 ACTCCTATTATGGAACAATCTCTCACTAAAAAGTTTCGGACTAAAGAAAGATGGT 237
Db 59 AspSerLysAsn-----ThrGlnThrValPheGlyLeuLysLysAspTrp 73
QY 238 GATATAACAAAAAAGACGAT-----TTTACAAGAGTA 270
Db 74 AspGlyValLysThrProSerAspSerSerAsnThrAsnSerThrIlePheThrGluLys 93
QY 271 GCTCAGCGCATGATTTTCAAATACTTAATPACAGGATTTTCAGGAAGTATTCGTTAC 330
Db 94 AspTyrSerPheArgTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyr 113
QY 331 TCTATGGACGACCAAGATAGAACTTGAAGTGCATATCAACAATTAATCAAAAAAC 390
Db 114 SerMetAsnGlyProArgIleGluPheGluValSerTyrGluThrPheAspValLysAsn 133
QY 391 ACCGATAACAATGATACATGATAATGGTGAATACATAAATTTTGCATTATCTCGTAAA 450
Db 134 LeuGlyGlyAsnTyrLysAsnAsnAlaHisMetTyrCysAlaLeuAspThrAlaAlaGln 153
QY 451 GATGCAATGGAAGAT-----CAGCAATATGTAGTACTTTAAATAATGACGGCAATACT 501
Db 154 AsnSerThrAsnGlyAlaGlyLeuThrThrSerValMetValLysAsnGluAsnLeuThr 173

; Sequence 44, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 44
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis p28-3 protein
US-09-660-587-44

Alignment Scores:
Pred. No.: 1.46e-51 Length: 276
Score: 520.00 Matches: 121
Percent Similarity: 57.93% Conservative: 47
Best Local Similarity: 41.72% Mismatches: 98
Query Match: 35.11% Indels: 24
DB: 4 Gaps: 8

US-10-062-624-41 (1-840) x US-09-660-587-44 (1-276)
QY 1 ATGAATTATAGAAAATCTAGTAGAGCGGTTAAATCTAATTAATCTAATCTTACCA 60
Db 1 MetAsnCysLysLeuValPheThrSerLeuValSerLeuIleThrThrLeuValSerLeuIleLeuLeuPro 20
QY 61 TATCAGTCTTTGAGATCTGTAGGTTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120
Db 21 GlyIleSerPheSerLysProIle-----HisGluAsnAsnThrThrGlyAsnPheTyr 38
QY 121 ATTAGTCAAGAGTACAATCCAAAGTATATCACACTTTAGAAAATCTCTGCTGAAGAACT 180
Db 39 IleIleGlyLysTyrValProSerIleSerHisPheGlyAsnPheSerAlaLysGlu---- 57
QY 181 CCTATTATGGAACAATAATCTCTCACTAAAAAGTTTTCGCACTAAAGAAAGAT----- 234
Db 58 -----GluLysAsnThrThrThrGlyIlePheGlyLeuLysGluSerTrpThr 73

; Sequence 44, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 44
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis p28-3 protein
US-09-660-587-44

Alignment Scores:
Pred. No.: 1.46e-51 Length: 276
Score: 520.00 Matches: 121
Percent Similarity: 57.93% Conservative: 47
Best Local Similarity: 41.72% Mismatches: 98
Query Match: 35.11% Indels: 24
DB: 4 Gaps: 8

US-10-062-624-41 (1-840) x US-09-660-587-44 (1-276)
QY 1 ATGAATTATAGAAAATCTAGTAGAGCGGTTAAATCTAATTAATCTAATCTTACCA 60
Db 1 MetAsnCysLysLeuValPheThrSerLeuValSerLeuIleThrThrLeuValSerLeuIleLeuLeuPro 20
QY 61 TATCAGTCTTTGAGATCTGTAGGTTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120
Db 21 GlyIleSerPheSerLysProIle-----HisGluAsnAsnThrThrGlyAsnPheTyr 38
QY 121 ATTAGTCAAGAGTACAATCCAAAGTATATCACACTTTAGAAAATCTCTGCTGAAGAACT 180
Db 39 IleIleGlyLysTyrValProSerIleSerHisPheGlyAsnPheSerAlaLysGlu---- 57
QY 181 CCTATTATGGAACAATAATCTCTCACTAAAAAGTTTTCGCACTAAAGAAAGAT----- 234
Db 58 -----GluLysAsnThrThrThrGlyIlePheGlyLeuLysGluSerTrpThr 73
```

```

Qy 235 ---GGTGATATAACAAAAAGACGATTTTACAGAGTAGTCTCCAGGCAATGATTTT--- 288
Db 74 GlyGlyIleIleLeuAspLysGluHisAlaAlaPheAenIleProAsnTyrSerPheLys 93
Qy 289 ---CAAAATACTTAATACAGGATTTTCAGGAGTATTGGTACTCTCTATGCGGACCA 345
Db 94 TyrGluAenAenProPheLeuGlyPheAlaGlyValIleGlyTyrSerIleGlySerPro 113
Qy 346 AGAATAGAACCTTGAAGCTGCATATCAACAAATTTAATCCAAAAAACACCCAGTAAACAATGAT 405
Db 114 ArgIleGluPheGluValSerTyrGluThrPheAspValGlnAenPro-----GlyAsp 131
Qy 406 ACTGATAATGGTGAATACATAACATTTTCATATCTCTGTAAGATGCA-----ATG 459
Db 132 LysPheAenAenAspAlaHisLysTyrCysAlaLeuSerAenAspSerSerLysThrMet 151
Qy 460 GAAGATCAGCAATATGATAGTACTTAAATAATCAGCGCATATCTTTATGTCATGATGGTT 519
Db 152 LysSerGlyLysPheValPheLeuLysAsnGluGlyLeuSerAspIleSerLeuMetLeu 171
Qy 520 AATACCTGCTATGATACATACAGCTGAAGGAGTATCTTTTCGACCATATGCAATGTCAGGT 579
Db 172 AsnValCysTyrAspIleIleAsnLysArgMetProPheSerProTyrIleCysAlaGly 191
Qy 580 ATAGAGCAGATCTTATCATTATTTTAAAGACCTCAATCTAATAATTTGCTTACCAAGGA 639
Db 192 IleGlyThrAspLeuIlePheMetPheAspAlaIleAenHisLysAlaAlaTyrGlnGly 211
Qy 640 AAAATAGTATAGTATACCTTATCACACCAAGAGTCTCTGATTTATTTGGTGGATCTAC 699
Db 212 LysLeuGlyPheAenTyrProIleSerProGluAlaAenIleSerMetGlyValHisPhe 231
Qy 700 CATGCGCTATTGGTAATAATTTGAGAAGATACCTGTAATAACTCTCTGTAGTATTAAT 759
Db 232 HisLysValThrAsnAenGluPheArg-----ValProValLeuLeuThr 246
Qy 760 GAT-----GCTCTCAAAACCAACATCTCTCTGATGATCTTGGAGTATCTTT 810
Db 247 AlaGlyGlyLeuAlaProAspAenLeuPheAlaIleValLysLeuSerIleCysHisPhe 266
Qy 811 GGCGAGAAATGGATGAGGTTACCTTC 840
Db 267 GlyLeuGluPheGlyTyrArgValSerPhe 276

```

RESULT 37

```

US-08-953-326-19
; Sequence 19, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UP-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-08-953-326-19

```

```

Alignment Scores:
Pred. No.: 4.9e-51 Length: 287
Score: 515.50 Matches: 120
Percent Similarity: 54.75% Conservative: 47
Best Local Similarity: 39.34% Mismatches: 95
Query Match: 34.81% Indels: 43
DB: Gaps: 7

US-10-062-624-41 (1-840) x US-08-953-326-19 (1-287)

Qy 1 ATGATTAATTAAGAAATCTAGTAAGAGCGCGTTAATCTCAATTAATGTCATCTTACCA 60
Db 1 MetLysTyrLysLysThrPheThrValThrAlaLeuValLeuLeuThrSerPheThrHis 20
Qy 61 TATCAGTCTTTTTCAGATCTCTAGGTTCAGAACTAATGATACAAAGAGGCTTCTTAC 120
Db 21 PheIleProPheTyrSerProAlaArgAlaSerThrIleHisAen-----PheTyr 37
Qy 121 ATTAGTCAAGTACAATCAAGPATATCACACTTTAGAAAATTTCTCTGTGAAAGAACT 180
Db 38 IleSerGlyLysTyrMetProThrAlaSerHisPheGlyIlePheSerAlaLysGlu--- 56
Qy 181 CCTATTATGGAACAATTTCTCTCACTAAAAGTTTTCGACTA-----AAGAAAGAT 234
Db 57 -----GluGlnSerPheThrLysValLeuValGlyLeuAspGlnArgLeuSer 72
Qy 235 GGTGATATAACAAAAAGACGATTTTACAGAGTAGCTCCA-----GGCATT 282
Db 73 HisAenIleIleAenAenAenAenPheThrAlaLysSerLeuLysValGlnAenTyrSerPhe 92
Qy 283 GATTTTCAAAATAACTTAATATCAGGATTTTTCAGGAAAGTATTTGTTACTCTCTATGCGG 342
Db 93 LysTyrLysAenAenProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerIleGlyAen 112
Qy 343 CCAGATAGACTTGAAGCTGCATATCAACATTTTAAATCCAAAAAACACCGATAACAT 402
Db 113 SerArgIleGluLeuGluValSerHisGluIlePheAspThrLysAenProGlyAenAen 132
Qy 403 -----GAT 405
Db 133 TyrLeuAenAenSerHisLysTyrCysAlaLeuSerHisGlySerHisIleCysSerAsp 152
Qy 406 ACTGATAATGGTGAATACTATAAACAATTTTGCATTTATCTCTGTAAGATGCAATGGAAGAT 465
Db 153 GlyAenSerGlyAspTrpTyr-----ThrAlaLysThr 163
Qy 466 CAGCAATATGTAGTACTTTAAAAATGACGCGCATAACTTTTATGTCATTTGTTAACTACT 525
Db 164 AspLysPheValLeuLeuLysAenGluGlyLeuLeuAspValSerPheMetLeuAenAla 183
Qy 526 TGCTATGACATACAGCTGAAGGAGTATCTTCTGATCATATGCTGTCAGGATATAGGA 585
Db 184 CysTyrAspIleThrThrGluLysMetProPheSerProTyrIleCysAlaGlyIleGly 203
Qy 586 GCAGATCTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGAAAAATA 645
Db 204 ThrAspLeuIleSerMetPheGluThrThrGlnAenLysIleSerTyrGlnGlyLysLeu 223
Qy 646 GGTATTAGTACCTATCAGCAAGATCTCTGCAATTTTATGTTGATGATACCATGGC 705
Db 224 GlyLeuAenTyrThrIleAenSerArgValSerValPheAlaGlyGlyHisPheHisLys 243
Qy 706 GTTATGCTTAATAATTTGAGAAGATACCTGTAATACTCTCTGATGATTTAAATGATGCT 765
Db 244 ValIleGlyAenGluPheLysGlyIleProThrLeuLeuProAspGlySerAenIleLys 263
Qy 766 CCTCAAAACCAATCTGCTCAGTAACCTCTTGACCTTGGATCTTTGGCGGAGAAATGGA 825
Db 264 ValGln---GlnSerAlaThrValThrLeuAspValCysHisPheGlyLeuGluIleGly 282
Qy 826 ATGAGGTTACCTTC 840

```

```
Db      283 SerArgPhePhe 287
RESULT 38
US-09-660-587-8
; Sequence 8, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 8
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of p28-4 protien
US-09-660-587-8

Alignment Scores:
Pred. No.: 4.9e-51 Length: 287
Score: 515.50 Matches: 120
Percent Similarity: 54.75% Conservative: 47
Best Local Similarity: 39.34% Mismatches: 95
Query Match: 34.81% Indels: 43
DB: 7

US-10-062-624-41 (1-840) x US-09-660-587-8 (1-287)
Qy      1 ATGATTATAGAAATCTAGTAGAGCGGTTAACTCATTATGCTCAATCTTACCA 60
Db      1 MetLysTyrLysLysThrPheThrValThrAlaLeuValLeuLeuThrSerPheThrHis 20
Qy      61 TATCAGTCTTTGGCAGATCCTGTAGGTTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120
Db      21 PheIleProPheTyrSerProAlaArgAlaSerThrIleHisAsn-----PheTyr 37
Qy      121 ATTAGTCAAGATGACAAATCCAAAGTATATCAACATTTAGAAAAATCTCTGCTGAAGAACT 180
Db      38 IleSerGlyLysTyrMetProThrAlaSerHisPheGlyIlePheSerAlaLysGlu--- 56
Qy      181 CCTATTATGGAACAAATCTCTCACTAAAGTTTTCGGACTA-----AAGAAAGAT 234
Db      57 -----GluGlnSerPheThrLysValLeuValGlyLeuAspGlnArgLeuSer 72
Qy      235 GGTGATATACAAAAAGACGATTTTTACAAGAGTAGCTCCA-----GGCATT 282
Db      73 HisAsnIleLeuAsnAsnAspThrAlaLysSerLeuLysValGlnAsnTyrSerPhe 92
Qy      283 GATTTCAAAATAACTTAATATCAGATTTTCAGGAAGTATTGGTTACTCTATGGAACGGA 342
Db      93 LysTyrLysAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerIleGlyAsn 112
Qy      343 CCAAGATAGACTTTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGATACAAAT 402
Db      113 SerArgIleGluLeuGluValSerHisGluIlePheAspThrLysAsnProGlyAsnAsn 132
Qy      403 -----GAT 405
Db      133 TyrLeuAsnAspSerHisLysTyrCysAlaLeuSerHisGlySerHisIleCysSerAsp 152
Qy      406 ACTGATATGGAATGATATATAAATTTTGGCATTATCTCGTAAGATGCAATGGAAGAT 465
Db      153 GlyAsnSerGlyAspTrpTyr-----ThrAlaLysThr 163
Qy      466 CAGCAATATGTAGTACTTAAATAAGACGCGCATAACTTTTATGCTCATTTGATGTTAATACT 525
```

```
Db      164 AspLysPheValLeuLeuLysAsnGluGlyLeuLeuAspValSerPheMetLeuAsnAla 183
Qy      526 TGCTATGACATTACAGCTGAAGAGTAGTATCTTTTCGTACATATGTCATGTCAGGTATAGGA 585
Db      184 CysTyrAspIleThrThrGluLysMetProPheSerProTyrIleCysAlaGlyIleGly 203
Qy      586 GCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAATA 645
Db      204 ThrAspLeuIleSerMetPheGluThrThrGlnAsnLysIleSerTyrGlnGlyLysLeu 223
Qy      646 GGTATTAGTTACCTCATCACACCAAGAGTCTCTGCATTATTGGTGGATATACCATCCGC 705
Db      224 GlyLeuAsnTyrThrIleAsnSerArgValSerValPheAlaGlyGlyHisPheHisLys 243
Qy      706 GTTATTGGTAATAAATTGGAAGATACCTGTGAATAACTCTCTGTAGTATTAAATGATGCT 765
Db      244 ValIleGlyAsnGluPheLysGlyIleProThrLeuLeuProAspGlySerAsnIleLys 263
Qy      766 CCTCAAAACCATCTGCTTCAGTAACCTTTCAGCTTGGATACTTTGGCGGAGAAATTCGA 825
Db      264 ValGln---GlnSerAlaThrValThrLeuAspValCysHisPheGlyLeuGluIleGly 282
Qy      826 ATGAGTTTCACCTTC 840
Db      283 SerArgPhePhe 287

RESULT 39
US-09-261-358A-8
; Sequence 8, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 8
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of ECa28SA1 protien
US-09-261-358A-8

Alignment Scores:
Pred. No.: 4.9e-51 Length: 287
Score: 515.50 Matches: 120
Percent Similarity: 54.75% Conservative: 47
Best Local Similarity: 39.34% Mismatches: 95
Query Match: 34.81% Indels: 43
DB: 7

US-10-062-624-41 (1-840) x US-09-261-358A-8 (1-287)
Qy      1 ATGATTATAGAAATCTAGTAGAGCGGTTAACTCATTATGCTCAATCTTACCA 60
Db      1 MetLysTyrLysLysThrPheThrValThrAlaLeuValLeuLeuThrSerPheThrHis 20
Qy      61 TATCAGTCTTTGGCAGATCCTGTAGGTTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120
Db      21 PheIleProPheTyrSerProAlaArgAlaSerThrIleHisAsn-----PheTyr 37
Qy      121 ATTAGTCAAGATGACAAATCCAAAGTATATCAACATTTAGAAAAATCTCTGCTGAAGAACT 180
Db      38 IleSerGlyLysTyrMetProThrAlaSerHisPheGlyIlePheSerAlaLysGlu--- 56
Qy      181 CCTATTATGGAACAAATCTCTCACTAAAGTTTTCGGACTA-----AAGAAAGAT 234
Db      57 -----GluGlnSerPheThrLysValLeuValGlyLeuAspGlnArgLeuSer 72
Qy      235 GGTGATATACAAAAAGACGATTTTTACAAGAGTAGCTCCA-----GGCATT 282
Db      73 HisAsnIleLeuAsnAsnAspThrAlaLysSerLeuLysValGlnAsnTyrSerPhe 92
Qy      283 GATTTCAAAATAACTTAATATCAGATTTTCAGGAAGTATTGGTTACTCTATGGAACGGA 342
Db      93 LysTyrLysAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerIleGlyAsn 112
Qy      343 CCAAGATAGACTTTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGATACAAAT 402
Db      113 SerArgIleGluLeuGluValSerHisGluIlePheAspThrLysAsnProGlyAsnAsn 132
Qy      403 -----GAT 405
Db      133 TyrLeuAsnAspSerHisLysTyrCysAlaLeuSerHisGlySerHisIleCysSerAsp 152
Qy      406 ACTGATATGGAATGATATATAAATTTTGGCATTATCTCGTAAGATGCAATGGAAGAT 465
Db      153 GlyAsnSerGlyAspTrpTyr-----ThrAlaLysThr 163
Qy      466 CAGCAATATGTAGTACTTAAATAAGACGCGCATAACTTTTATGCTCATTTGATGTTAATACT 525
```



Db 283 SerArgPhePhe 287  
|||||

Search completed: July 8, 2003, 10:50:48  
Job time : 35 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 8, 2003, 10:48:37 ; Search time 51.5 Seconds  
(without alignments)

3797.771 Million cell updates/sec

Title: US-10-062-624-41

Perfect score: 1481

Sequence: 1 atgaattataagaaattctt.....ttggaatgaggttcaccttc 840

Scoring table:

	BLOSUM62	Xgapop 10.0	Xgapext 0.5
Xgapop 10.0			0.5
Xgapext 0.5			
Fgapop 6.0			7.0
Delopt 6.0			

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 891516

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool/US10062624/runat\_08072003\_093522\_2182/app\_query.fasta\_1.1031  
-DB=Published Applications AA-QPWT=fastan -SUFFIX=rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=Dice -START=1 -END=1 -MATRIX=Dblsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10062624@cgn 1.17 @runat\_08072003\_093522\_2182  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGESQUERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRSADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pap.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pap.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pap.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pap.\*  
5: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pap.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pap.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pap.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pap.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pap.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pap.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pap.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pap.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pap.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1462	98.7	280	9	US-10-062-051-42 Sequence 42, Appl
2	1462	98.7	280	9	US-10-062-920-42 Sequence 42, Appl
3	1462	98.7	280	9	US-10-314-639-48 Sequence 48, Appl
4	1462	98.7	280	12	US-10-062-624-42 Sequence 42, Appl

5	1462	98.7	280	12	US-10-059-964-48	Sequence 48, Appl
6	1202.5	81.2	283	9	US-10-062-051-10	Sequence 10, Appl
7	1202.5	81.2	283	9	US-10-284-986-14	Sequence 14, Appl
8	1202.5	81.2	283	9	US-10-062-920-10	Sequence 10, Appl
9	1202.5	81.2	283	9	US-10-314-639-4	Sequence 4, Appl
10	1202.5	81.2	283	10	US-09-846-808-14	Sequence 14, Appl
11	1202.5	81.2	283	12	US-10-062-624-10	Sequence 10, Appl
12	1202.5	81.2	283	12	US-10-059-964-4	Sequence 4, Appl
13	644.5	43.5	281	9	US-10-062-051-9	Sequence 9, Appl
14	644.5	43.5	281	9	US-10-284-986-19	Sequence 19, Appl
15	644.5	43.5	281	9	US-10-062-920-9	Sequence 9, Appl
16	644.5	43.5	281	10	US-09-846-808-19	Sequence 19, Appl
17	644.5	43.5	281	12	US-10-062-624-9	Sequence 9, Appl
18	642.5	43.4	281	9	US-10-314-639-2	Sequence 2, Appl
19	642.5	43.4	281	12	US-10-059-964-2	Sequence 18, Appl
20	629.5	42.5	276	12	US-10-062-994-18	Sequence 18, Appl
21	629.5	42.5	276	12	US-10-062-994-18	Sequence 32, Appl
22	621	41.9	288	9	US-10-314-639-32	Sequence 32, Appl
23	621	41.9	288	12	US-10-059-964-32	Sequence 32, Appl
24	620	41.9	286	9	US-10-062-051-12	Sequence 12, Appl
25	620	41.9	286	9	US-10-284-986-16	Sequence 16, Appl
26	620	41.9	286	9	US-10-062-920-12	Sequence 12, Appl
27	620	41.9	286	9	US-10-314-639-8	Sequence 8, Appl
28	620	41.9	286	10	US-09-846-808-16	Sequence 16, Appl
29	620	41.9	286	12	US-10-062-994-15	Sequence 15, Appl
30	620	41.9	286	12	US-10-062-624-12	Sequence 12, Appl
31	620	41.9	286	12	US-10-059-964-8	Sequence 8, Appl
32	620	41.9	286	12	US-10-062-994-15	Sequence 15, Appl
33	605	40.9	280	9	US-10-062-051-14	Sequence 14, Appl
34	605	40.9	280	9	US-10-062-920-14	Sequence 14, Appl
35	605	40.9	280	12	US-10-314-639-12	Sequence 12, Appl
36	605	40.9	280	12	US-10-062-624-14	Sequence 14, Appl
37	605	40.9	280	12	US-10-059-964-12	Sequence 12, Appl
38	603	40.7	280	9	US-10-284-986-18	Sequence 18, Appl
39	603	40.7	280	10	US-09-846-808-18	Sequence 18, Appl
40	603	40.7	280	12	US-10-062-994-17	Sequence 17, Appl
41	603	40.7	280	12	US-10-062-994-17	Sequence 17, Appl
42	601	40.6	280	9	US-10-062-051-6	Sequence 6, Appl
43	601	40.6	280	9	US-10-062-920-6	Sequence 6, Appl
44	601	40.6	280	9	US-10-314-639-38	Sequence 38, Appl
45	601	40.6	280	12	US-10-062-624-6	Sequence 6, Appl

#### ALIGNMENTS

RESULT 1  
US-10-062-051-42  
; Sequence 42, Application US/10062051  
; Publication No. US20030073095A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: US/10/062,051  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: US/09/660,587  
; PRIOR FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 42  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of E. canis p28-2 protein  
US-10-062-051-42

Alignment Scores: 7.86e-150 Length: 280  
Pred. No.: 280

Score: 1462.00 Matches: 280  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 98.72% Indels: 0  
DB: 9 Gaps: 0

US-10-062-624-41 (1-840) x US-10-062-051-42 (1-280)

Qy 1 ATGAATTAAGAAATCTAGTAAGAGCGGTTAATCTCAATTAATGCTCAATCTTACCA 60  
Db 1 MetAenTyrlsYsIleLeuValArgSerAlaLeuIleSerLeuMetSerIleLeuPro 20

Qy 61 TATCAGTCTTTGCGAGATCTGTAGTTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120  
Db 21 TyrGlnSerPheAlaAspProValGlySerArgThrAsnAspAsnLysGluGlyPheTyr 40

Qy 121 ATTAGTCAAAAGTACAATCCAAAGTATATCACACTTTAGAAAATTTCTGCTGCAAGAACT 180  
Db 41 IleSerAlaLysTyrlsYsIleLeuValArgSerAlaLeuIleSerLeuMetSerIleLeuPro 60

Qy 181 CCTATTATGGAACAAATTTCTCTCACTAAAAGTCTTCAAGAACTTCTGCTGCAAGAACT 240  
Db 61 ProIleAsnGlyThrAsnSerLeuThrLysLysValPheGlyLeuLysLysAspGlyAsp 80

Qy 241 ATACAAAAAAGACGATTTTACAGAGTAGCTCCAGGCAATTTGATTTTCAAAATAACTTA 300  
Db 81 IleThrLysLysAspPheThrArgValAlaProGlyIleAspPheGlnAsnAsnLeu 100

Qy 301 ATATCAGGATTTTACAGGAGTATTTGGTTACTCTCTGAGCGGACCAAGAACTTGAAT 360  
Db 101 IleSerGlyPheSerGlySerIleGlyTyrSerMetAspGlyProArgIleGluLeuGlu 120

Qy 361 GCTGCATATCAACAAATTTAATCCAAAGTATATCACACTTTAGAAAATTTCTGCTGCAAGAACT 420  
Db 121 AlaAlaTyrGlnGlnPheAsnProLysAsnThrAspAsnAsnAspThrAspAsnGlyGlu 140

Qy 421 TACTATTAACATTTTGCATTTATTTGATGATGATGATGATGATGATGATGATGATGATGAT 480  
Db 141 TyrTyrLysHisPheAlaLeuSerArgLysAspAlaMetGluAspGlnGlnTyrValVal 160

Qy 481 CTAAAAATGACGCGCAATACTTTTATGTCATTTGATGATGATGATGATGATGATGATGATGAT 540  
Db 161 LeuLysAsnAspGlyIleThrPheMetSerLeuMetValAsnThrCysTyrAspIleThr 180

Qy 541 GCTGAGAGATCTTTTGCATTTATTTGATGATGATGATGATGATGATGATGATGATGATGAT 600  
Db 181 AlaGluGlyValSerPheValProTyrAlaCysAlaGlyIleGlyAlaAspLeuIleThr 200

Qy 601 ATTTTAAAGACCTCAATCTAAAATTTGCTTACCAAGGAAATAGGTTAGTTAGTTACCT 660  
Db 201 IlePheLysAspLeuAsnLeuLysPheAlaTyrGlnGlyLysIleGlyIleSerTyrPro 220

Qy 721 TTTGAGAGATCTCTGCAATTTATTTGATGATGATGATGATGATGATGATGATGATGATGAT 780  
Db 241 PheGluLysIleProValIleThrProValValLeuAsnAspAlaProGlnThrThrSer 260

Qy 781 GCTTCAGTAATCTTTGAGCTTGTGATCTTTGCGGAGAAATGGAATGAGGTTTACCTTC 840  
Db 261 AlaSerValThrLeuAspValGlyTyrPheGlyGlyGluIleGlyWecArgPheThrPhe 280

## RESULT 2

US-10-062-920-42  
; Sequence 42, Application US/10062920  
; Publication No. US20030096250A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein

; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: US/10/062,920  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: US/09/660,587  
; PRIOR FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 42  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of E. canis p28-2 protein  
US-10-062-920-42

Alignment Scores:  
Pred. No.: 7,86e-150 Length: 280  
Score: 1462.00 Matches: 280  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 98.72% Indels: 0  
DB: 9 Gaps: 0

US-10-062-624-41 (1-840) x US-10-062-920-42 (1-280)

Qy 1 ATGAATTAAGAAATTTAGTAAGAGCGGTTAATCTCAATTAATGCTCAATCTTACCA 60  
Db 1 MetAenTyrlsYsIleLeuValArgSerAlaLeuIleSerLeuMetSerIleLeuPro 20

Qy 61 TATCAGTCTTTGCGAGATCTGTAGTTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120  
Db 21 TyrGlnSerPheAlaAspProValGlySerArgThrAsnAspAsnLysGluGlyPheTyr 40

Qy 121 ATTAGTCAAAAGTACAATCCAAAGTATATCACACTTTAGAAAATTTCTGCTGCAAGAACT 180  
Db 41 IleSerAlaLysTyrlsYsIleLeuValArgSerAlaLeuIleSerLeuMetSerIleLeuPro 60

Qy 181 CCTATTATGGAACAAATTTCTCTCACTAAAAGTCTTCAAGAACTTCTGCTGCAAGAACT 240  
Db 61 ProIleAsnGlyThrAsnSerLeuThrLysLysValPheGlyLeuLysLysAspGlyAsp 80

Qy 241 ATACAAAAAAGACGATTTTACAGAGTAGCTCCAGGCAATTTGATTTTCAAAATAACTTA 300  
Db 81 IleThrLysLysAspPheThrArgValAlaProGlyIleAspPheGlnAsnAsnLeu 100

Qy 301 ATATCAGGATTTTACAGGAGTATTTGGTTACTCTCTGAGCGGACCAAGAACTTGAAT 360  
Db 101 IleSerGlyPheSerGlySerIleGlyTyrSerMetAspGlyProArgIleGluLeuGlu 120

Qy 361 GCTGCATATCAACAAATTTAATCCAAAGTATATCACACTTTAGAAAATTTCTGCTGCAAGAACT 420  
Db 121 AlaAlaTyrGlnGlnPheAsnProLysAsnThrAspAsnAsnAspThrAspAsnGlyGlu 140

Qy 421 TACTATTAACATTTTGCATTTATTTGATGATGATGATGATGATGATGATGATGATGATGAT 480  
Db 141 TyrTyrLysHisPheAlaLeuSerArgLysAspAlaMetGluAspGlnGlnTyrValVal 160

Qy 481 CTAAAAATGACGCGCAATACTTTTATGTCATTTGATGATGATGATGATGATGATGATGATGAT 540  
Db 161 LeuLysAsnAspGlyIleThrPheMetSerLeuMetValAsnThrCysTyrAspIleThr 180

Qy 541 GCTGAGAGATCTTTTGCATTTATTTGATGATGATGATGATGATGATGATGATGATGATGAT 600  
Db 181 AlaGluGlyValSerPheValProTyrAlaCysAlaGlyIleGlyAlaAspLeuIleThr 200

Qy 601 ATTTTAAAGACCTCAATCTAAAATTTGCTTACCAAGGAAATAGGTTAGTTAGTTACCT 660  
Db 201 IlePheLysAspLeuAsnLeuLysPheAlaTyrGlnGlyLysIleGlyIleSerTyrPro 220

Qy 721 TTTGAGAGATCTCTGCAATTTATTTGATGATGATGATGATGATGATGATGATGATGATGAT 780  
Db 241 PheGluLysIleProValIleThrProValValLeuAsnAspAlaProGlnThrThrSer 260

Qy 781 GCTTCAGTAATCTTTGAGCTTGTGATCTTTGCGGAGAAATGGAATGAGGTTTACCTTC 840  
Db 261 AlaSerValThrLeuAspValGlyTyrPheGlyGlyGluIleGlyWecArgPheThrPhe 720

Db 221 IleThrProGluValSerAlaPheIleGlyTyrTyrHisGlyValIleGlyAsnLys 240  
Qy 721 TTTGAGAGACTGTAATAACTCTCTAGTATTAAATGATGCTCTCAACACACATCT 780  
Db 241 PheGluLysIleProValIleThrProValValLeuAsnAspAlaProGlnThrThrSer 260  
Qy 781 GCTTCAGTAACTCTGAGTGGTACTTTGGCGAGAAATGGAAGAGTTTACCTTC 840  
Db 261 AlaSerValThrLeuAspValGlyTyrPheGlyGlyGluIleGlyMetArgPheThrPhe 280

## RESULT 3

US-10-314-639-48  
; Sequence 48, Application US/10314639  
; Publication No. US20030103991A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20030103991A1  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; TITLE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/314,639  
; CURRENT FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US/09/314,701  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 48  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
US-10-314-639-48

## Alignment Scores:

Pred. No.: 7,86e-150 Length: 280  
Score: 1462.00 Matches: 280  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 98.72% Indels: 0  
DB: 9 Gaps: 0

US-10-062-624-41 (1-840) x US-10-314-639-48 (1-280)

Qy 1 ATGAATTTAAGAAAATCTAGTAAGAGCGCTTAATCTCATTAATGTCATCTTACCA 60  
Db 1 MetAsnTyrLysIleLeuValArgSerAlaLeuIleSerLeuMetSerIleLeuPro 20  
Qy 61 TATCAGTCTTTTGCAGATCTGTAGTTCAGAACTAATGATACAAAGAGGCTTCTAC 120  
Db 21 TyrGlnSerPheAlaAspProValGlySerArgThrAsnAspAsnLysGluGlyPheTyr 40  
Qy 121 ATTAGTCAAGTACAATCCAAAGTATATCACATTTAGAAAATCTCTGCTGAAGAACT 180  
Db 41 IleserAlaLysTyrAsnProSerIleSerHisPheArgLysPheSerAlaGluGluThr 60  
Qy 181 CCTATTATGGAACAATTTCTCTCACTAAAAAGTTTTCGACTAAAGAAAGATGGTGAT 240  
Db 61 ProIleAsnGlyThrAsnSerLeuThrLysLysValPheGlyLeuLysLysAspGlyAsp 80  
Qy 241 ATACAAAAAAGACGATTTTACAAAGTAGCTCCAGGCATTCGATTTTCAAAATACTTA 300  
Db 81 IleThrLysLysAspAspPheThrArgValAlaProGlyIleAspPheGlnAsnAsnLeu 100  
Qy 301 ATATCAGGATTTTTCAGGAAGTATTGGTTACTCTATGACCGGACCAAGAATAGAACTTGA 360  
Db 101 IleserGlyPheSerGlySerIleGlyTyrSerMetAspGlyProArgIleGluLeuGlu 120  
Qy 361 GCTGCATATCAACATTTATCCAAAAACACCGATTAACATGATCTGATTAATGGTGAA 420  
Db 121 AlaAlaTyrGlnGlnPheAsnProLysAsnThrAspAsnAsnAspThrAspAsnGlyGlu 140  
Qy 421 TACTATAAATTTTGCATATCTCTCGTAAAGATGCATGCAAGATCAGCAATATGTAGTA 480  
Db 141 TyrTyrLysHisPheAlaLeuSerArgLysAspAlaMetGluAspGlnGlnTyrValVal 160

Qy 481 CTTAAAAATGACGCGCATAACTTTTATGTCTATGATGTTAATCTTCTATGACATTACA 540  
Db 161 LeuLysAsnAspGlyIleThrPheMetSerLeuMetValAsnThrCysTyrAspIleThr 180  
Qy 541 GCTGAAGAGTATCTTCGTACCATATGATGTCAGGTATAGGAGCAGATCTTATCACT 600  
Db 181 AlaGluGlyValSerPheValProTyrAlaCysAlaGlyIleGlyAlaAspLeuIleThr 200  
Qy 601 ATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGAAAATAGGTATTAGTTACCT 660  
Db 201 IlePheLysAspLeuAsnLeuLysPheAlaTyrGlnGlyLysIleGlyIleSerTyrPro 220  
Qy 661 ATCACACCAGAGTCTCTGCATTTATTGGTGGATACCTACCATGCGTTATTGGTAATAA 720  
Db 221 IleThrProGluValSerAlaPheIleGlyTyrTyrHisGlyValIleGlyAsnLys 240  
Qy 721 TTTGAGAGTACTCTGTAATAACTCTGTAGTATTAAATGATGCTCTCAACACATCT 780  
Db 241 PheGluLysIleProValIleThrProValValLeuAsnAspAlaProGlnThrThrSer 260  
Qy 781 GCTTCAGTAACTCTGACGTTGATACTTTGGCGGAGAAATTTGAATGAGGTTCCACCTTC 840  
Db 261 AlaSerValThrLeuAspValGlyTyrPheGlyGlyGluIleGlyMetArgPheThrPhe 280

## RESULT 4

US-10-062-624-42  
; Sequence 42, Application US/10062624  
; Patent No. US20020115840A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-Kilodalton Immunodominant Protein  
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof  
; FILE REFERENCE: D6152CIP2/DI  
; CURRENT APPLICATION NUMBER: US/10/062,624  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: 09/660,587  
; PRIOR FILING DATE: 2000-09-12  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 42  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of E. canis p28-2 protein  
US-10-062-624-42

## Alignment Scores:

Pred. No.: 7,86e-150 Length: 280  
Score: 1462.00 Matches: 280  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 98.72% Indels: 0  
DB: 12 Gaps: 0

US-10-062-624-41 (1-840) x US-10-062-624-42 (1-280)

Qy 1 ATGAATTTAAGAAAATCTAGTAAGAGCGCTTAATCTCATTAATGTCATCTTACCA 60  
Db 1 MetAsnTyrLysIleLeuValArgSerAlaLeuIleSerLeuMetSerIleLeuPro 20  
Qy 61 TATCAGTCTTTTGCAGATCTGTAGTTCAGAACTAATGATACAAAGAGGCTTCTAC 120  
Db 21 TyrGlnSerPheAlaAspProValGlySerArgThrAsnAspAsnLysGluGlyPheTyr 40  
Qy 121 ATTAGTCAAGTACAATCCAAAGTATATCACATTTAGAAAATCTCTGCTGAAGAACT 180  
Db 41 IleserAlaLysTyrAsnProSerIleSerHisPheArgLysPheSerAlaGluGluThr 60  
Qy 181 CTTATTATGGAACAATTTCTCTCACTAAAAAGTTTTCGACTAAAGAAAGATGGTGAT 240

Db 61 ProIleAsnGlyThrAsnSerLeuThrLysLysValPheGlyLeuLysLysAspGlyAsp 80  
Qy 241 ATACAAAAAGACGATTTCACAGAGTAGCTCCAGGCATTGATTTTCAAAATAACTTA 300  
Db 81 IleThrLysLysAspPheThrArgValAlaProGlyIleAspPheGlnAsnLeu 100  
Qy 301 ATATCAGATTTTCAGGAGTATGGTTACTCTATGACGGCCAGCAAGTAATAGACTTCAA 360  
Db 101 IleSerGlyPheSerGlySerIleGlyTyrSerMetAspGlyProArgIleGluLeu 120  
Qy 361 GCTGATATCAACAAATTTAAATCCAAAAACCCGATAACAATGATACATAATGGTGAA 420  
Db 121 AlaAlaTyrGlnGlnPheAsnProLysAsnThrAspAsnAspThrAspAsnGlyGlu 140  
Qy 421 TACTATAAACATTTTGCATTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTA 480  
Db 141 TyrTyrLysHiePheAlaLeuSerArgLysAspAlaMetGluAspGlnGlnTyrVal 160  
Qy 481 CTTAAAAATGACGCATACACTTTTATGTCATTGATGGTAACTCTGCTATGCATTAACA 540  
Db 161 LeuLysAsnAspGlyIleThrPheMetSerLeuMetValAsnThrCysTyrAspIleThr 180  
Qy 541 GCTGAAGCAGATCTTTTCGTACCATATGATGTCAGGTATAGGACGAGATCTTATCACT 600  
Db 181 AlaGluGlyValSerPheValProTyrAlaCysAlaGlyIleGlyAlaAspLeuIleThr 200  
Qy 601 ATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAATAGGTATTAGTTACCT 660  
Db 201 IlePheLysAspLeuAsnLeuLysPheAlaTyrGlnGlyLysIleGlyLysSerTyrPro 220  
Qy 661 ATCCACACAGAGTCTCTGCATTTATTTGCTGATACCTACCTACCTGCGTATTGGTAATAA 720  
Db 221 IleThrProGluValSerAlaPheIleGlyGlyTyrTyrHisGlyValIleGlyAsnLys 240  
Qy 721 TTTGAGAAGATACCTGTAATAACTCTCTAGTATTAAATGATGCTCCCTCAAAACCATCT 780  
Db 241 PheGluLysIleProValIleThrProValValLeuAsnAspAlaProGlnThrThrSer 260  
Qy 781 GCTTCAGTAACCTTTGAGCTGGATGATCTTTGGCGGAGAAATGGAAATGAGTTTCACTTC 840  
Db 261 AlaSerValThrLeuAspValGlyTyrPheGlyGlyGluIleGlyMetArgPheThrPhe 280

## RESULT 5

US-10-059-964-48

; Sequence 48, Application US/10059964

; Patent No. US20020120115A1

; GENERAL INFORMATION:

; APPLICANT: Rikihisa, Yasuko

; APPLICANT: Ohasi, No. US20020120115A1io

; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia

; FILE OF INVENTION: Chaffeeensis

; FILE REFERENCE: 22727/04021

; CURRENT APPLICATION NUMBER: US/10/059,964

; EARLIER FILING DATE: 2002-01-28

; EARLIER APPLICATION NUMBER: 09/314,701

; EARLIER FILING DATE: 1999-05-19

; NUMBER OF SEQ ID NOS: 66

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 48

; LENGTH: 280

; TYPE: PRT

; ORGANISM: Ehrlichia canis

US-10-059-964-48

## Alignment Scores:

Pred. No.:	7,86e-150	Length:	280
Score:	1462.00	Matches:	280
Percent Similarity:	100.00%	Conservatives:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	98.72%	Indels:	0
DB:	12	Gaps:	0

US-10-062-624-41 (1-840) x US-10-059-964-48 (1-280)

Qy 1 ATGNAATTATAAGRAAATTTCTAGTAAGAGCGCGTAAATCTCATTAATGTCAATCTTACCA 60  
Db 1 MetAsnTyrLysLysIleLeuValArgSerAlaLeuIleSerLeuMetSerIleLeuPro 20  
Qy 61 TATCAGTCTTTTTCAGATCTCTGAGGTTCAAGAACTAATGATACAAAGAGGCTTCTTAC 120  
Db 21 TyrGlnSerPheAlaAspProValGlySerArgThrAsnAspAsnLysGluGlyPheTyr 40  
Qy 121 ATTAGTCCAAAGTACAATCCAAAGTATATCACATTTTGAAGAAATTTCTCTGTAAGAAACT 180  
Db 41 IleSerAlaLysTyrAsnProSerIleSerHisPheArgLysPheSerAlaGluGluThr 60  
Qy 181 CCTATTAAATGGAACAAATTTCTCTCACTAAAAAGTTTTCGACCTAAAGAGATGGTGAT 240  
Db 61 ProIleAsnGlyThrAsnSerLeuThrLysLysValPheGlyLeuLysLysAspGlyAsp 80  
Qy 241 ATACAAAAAAGACGATTTTACAAAGTAGTCTCCAGGCATTGATTTTCAAAATAACTTA 300  
Db 81 IleThrLysLysAspAspPheThrArgValAlaProGlyIleAspPheGlnAsnAsnLeu 100  
Qy 301 ATATCAGATTTTTCAGGAGTATGGTTACTCTATGACGGCCAGCAAGTAATAGACTTCAA 360  
Db 101 IleSerGlyPheSerGlySerIleGlyTyrSerMetAspGlyProArgIleGluLeuGlu 120  
Qy 361 GCTGCATATCAACAAATTTTAAATCCAAAAACCCGATAACAATGATGATCTGATTAATGGTCAA 420  
Db 121 AlaAlaTyrGlnGlnPheAsnProLysAsnThrAspAsnAsnAspThrAspAsnGlyGlu 140  
Qy 421 TACTATAAACATTTTGCATTATCTCTGTAAGATGCAATGGAAGATCAGCAATATGTA 480  
Db 141 TyrTyrLysHisPheAlaLeuSerArgLysAspAlaMetGluAspGlnGlnTyrValVal 160  
Qy 481 CTTAAAAATGACGCATACACTTTTATGTCATTGATGGTAACTCTGCTATGCATTAACA 540  
Db 161 LeuLysAsnAspGlyIleThrPheMetSerLeuMetValAsnThrCysTyrAspIleThr 180  
Qy 541 GCTGAAGCAGTATCTTTTCGTACCATATGATGTCAGGTATAGGACGAGATCTTATCACT 600  
Db 181 AlaGluGlyValSerPheValProTyrAlaCysAlaGlyIleGlyAlaAspLeuIleThr 200  
Qy 601 ATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAATAGGTATTAGTTACCT 660  
Db 201 IlePheLysAspLeuAsnLeuLysPheAlaTyrGlnGlyLysIleGlyLysSerTyrPro 220  
Qy 661 ATCCACACAGAGTCTCTGCATTTATTTGCTGATACCTACCTACCTGCGTATTGGTAATAA 720  
Db 221 IleThrProGluValSerAlaPheIleGlyGlyTyrTyrHisGlyValIleGlyAsnLys 240  
Qy 721 TTTGAGAAGATACCTGTAATAACTCTCTAGTATTAAATGATGCTCCCTCAAAACCATCT 780  
Db 241 PheGluLysIleProValIleThrProValValLeuAsnAspAlaProGlnThrThrSer 260  
Qy 781 GCTTCAGTAACCTTTGAGCTGGATGATCTTTGGCGGAGAAATGGAAATGAGTTTCACTTC 840  
Db 261 AlaSerValThrLeuAspValGlyTyrPheGlyGlyGluIleGlyMetArgPheThrPhe 280

## RESULT 6

US-10-062-051-10

; Sequence 10, Application US/10062051

; Publication No. US20030073095A1

; GENERAL INFORMATION:

; APPLICANT: Walker, David H.

; APPLICANT: McBride, Jere W.

; APPLICANT: Yu, Xue-Jie

; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein

; FILE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof

; FILE REFERENCE: D6152CIP2

; CURRENT APPLICATION NUMBER: US/10/062,051

; CURRENT FILING DATE: 2002-01-31

; PRIOR APPLICATION NUMBER: US/09/660,587

; PRIOR FILING DATE: 2000-09-12

; PRIOR APPLICATION NUMBER: 09/261,358

```
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 10
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1B
US-10-062-051-10

Alignment Scores:
Pred. No.: 1,08e-121 Length: 283
Score: 1202.50 Matches: 225
Percent Similarity: 88.69% Conservative: 26
Best Local Similarity: 79.51% Mismatches: 29
Query Match: 81.20% Indels: 3
DB: 9 Gaps: 1

US-10-062-624-41 (1-840) x US-10-062-051-10 (1-283)
Qy 1 ATGAATTATAGAAAATCTAGTAAGAGCGGTTAATCTCATTAAATGTCATCTTACCA 60
Db 1 MetAsnTyrLysLysIlePheValSerSerAlaLeuIleSerLeuMetSerIleLeuPro 20
Qy 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAAGA-----ACTAATGATAACAAGAA 111
Db 21 TyGlnSerPheAlaAspProValThrSerAsnAspThrGlyIleAsnAspSerArgGlu 40
Qy 112 GGCTTCTACATTAGTGCAGATCCTGTAGTTCAAGA-----ACTAATGATAACAAGAA 171
Db 41 GlyPheTyrIleSerValLysTyrAsnProSerIleSerHisPheArgLysPheSerAla 60
Qy 172 GAAGAACTCTTATTAATGGAACAATCTCTCACTAAAAAGTTTTCGGACTAAAGAAA 231
Db 61 GluGluAlaProIleAsnGlyAsnThrSerIleThrLysLysValPheGlyLeuLysLys 80
Qy 232 GATGGTGATATAACAAGAGACGATTTTACAGAGTAGCTCCAGGCAATGATTTCAA 291
Db 81 AspGlyAspIleAlaGlnSerAlaAsnPheAsnArgThrAspProAlaLeuGluPheGln 100
Qy 292 AATAACTTAAATGAGAGTATCTTGCATCATATGCTGATGATGATGATGATGATGAT 351
Db 101 AsnAsnLeuIleSerGlyPheSerGlySerIleGlyTyrAlaMetAspGlyProArgIle 120
Qy 352 GAACCTGAGTGAAGAGTATCTTGCATCATATGCTGATGATGATGATGATGATGAT 411
Db 121 GluLeuGluAlaAlaTyrGlnLysPheAspAlaLysAsnProAspAsnAsnAspThrAsn 140
Qy 412 AATGGTGATATAACAAGAGTATCTTGCATCATATGCTGATGATGATGATGATGAT 471
Db 141 SerGlyAspTyrTyrLysTyrPheGlyLeuSerArgGluAspAlaIleAlaAspLysLys 160
Qy 472 TATGTAGTACTTAAATGAGAGTATCTTGCATCATATGCTGATGATGATGATGATGAT 531
Db 161 TyrValValLeuLysAsnGluGlyIleThrPheMetSerLeuMetValAsnThrCysTyr 180
Qy 532 GACATTACAGTGAAGAGTATCTTGCATCATATGCTGATGATGATGATGATGATGAT 591
Db 181 AspIleThrAlaGluGlyValProPheIleProTyrAlaCysAlaGlyValGlyAlaAsp 200
Qy 592 CTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAATAGGTATT 651
Db 201 LeuIleAsnValPheLysAspPheAsnLeuLysPheSerTyrGlnGlyLysIleGlyIle 220
Qy 652 AGTTACCTATCACACAGAGTCTCTGCAATTTATGCTGATGATGATGATGATGATGAT 711
Db 221 SerTyrProIleThrProGluValSerAlaPheIleGlyTyrTyrHisGlyValIle 240
Qy 712 GGTAATAAATTTGAGAAGATACCTGTAATACTCTCTGATGATGATGATGATGATGAT 771
Db 241 GlyAsnAsnPheAsnLysIleProValIleThrProValValLeuGluGlyAlaProGln 260
Qy 772 ACCAATCTGCTTCAGTAACCTCTTACGCTTGATGATGATGATGATGATGATGATGAT 831
```

---

```
Db 261 ThrThrSerAlaLeuValThrIleAspThrGlyTyrPheGlyGlyGluValGlyValArg 280
Qy 832 TTCACCTTC 840
Db 281 PheThrPhe 283

RESULT 7
US-10-284-986-14
; Sequence 14, Application US/10284986
; Publication No. US20030091588A1
; GENERAL INFORMATION: David H.
; APPLICANT: Walker, David H.
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane
; TITLE OF INVENTION: Protein Multigene Family
; FILE REFERENCE: D6311D1
; CURRENT APPLICATION NUMBER: US/10/284,986
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: 09/846,808
; PRIOR FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 14
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: P28-14 Outer Membrane Protein of
; OTHER INFORMATION: Ehrlichia chaffeensis
US-10-284-986-14

Alignment Scores:
Pred. No.: 1,08e-121 Length: 283
Score: 1202.50 Matches: 225
Percent Similarity: 88.69% Conservative: 26
Best Local Similarity: 79.51% Mismatches: 29
Query Match: 81.20% Indels: 3
DB: 9 Gaps: 1

US-10-062-624-41 (1-840) x US-10-284-986-14 (1-283)
Qy 1 ATGAATTATAGAAAATCTAGTAAGAGCGGTTAATCTCATTAAATGTCATCTTACCA 60
Db 1 MetAsnTyrLysLysIlePheValSerSerAlaLeuIleSerLeuMetSerIleLeuPro 20
Qy 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAAGA-----ACTAATGATAACAAGAA 111
Db 21 TyGlnSerPheAlaAspProValThrSerAsnAspThrGlyIleAsnAspSerArgGlu 40
Qy 112 GGCTTCTACATTAGTGCAGATCCTGTAGTTCAAGA-----ACTAATGATAACAAGAA 171
Db 41 GlyPheTyrIleSerValLysTyrAsnProSerIleSerHisPheArgLysPheSerAla 60
Qy 172 GAAGAACTCTTATTAATGGAACAATCTCTCACTAAAAAGTTTTCGGACTAAAGAAA 231
Db 61 GluGluAlaProIleAsnGlyAsnThrSerIleThrLysLysValPheGlyLeuLysLys 80
Qy 232 GATGGTGATATAACAAGAGACGATTTTACAGAGTAGCTCCAGGCAATGATTTCAA 291
Db 81 AspGlyAspIleAlaGlnSerAlaAsnPheAsnArgThrAspProAlaLeuGluPheGln 100
Qy 292 AATAACTTAAATGAGAGTATCTTGCATCATATGCTGATGATGATGATGATGATGAT 351
Db 101 AsnAsnLeuIleSerGlyPheSerGlySerIleGlyTyrAlaMetAspGlyProArgIle 120
Qy 352 GAACCTGAGTGAAGAGTATCTTGCATCATATGCTGATGATGATGATGATGATGAT 411
Db 121 GluLeuGluAlaAlaTyrGlnLysPheAspAlaLysAsnProAspAsnAsnAspThrAsn 140
Qy 412 AATGGTGATATAACAAGAGTATCTTGCATCATATGCTGATGATGATGATGATGAT 471
Db 141 SerGlyAspTyrTyrLysTyrPheGlyLeuSerArgGluAspAlaIleAlaAspLysLys 160
```



## Alignment Scores:

Pred. No.: 1,08e-121 Length: 283  
Score: 1202.50 Matches: 225  
Percent Similarity: 88.69% Conservative: 26  
Best Local Similarity: 79.51% Mismatches: 29  
Query Match: 81.20% Indels: 3  
DB: 9 Gaps: 1

US-10-062-624-41 (1-840) x US-10-314-639-4 (1-283)

Qy 1 ATGAATTATAAGAAAATTCAGTAAGAGCGCGTTAATCTCATTAATGTCAATCTTACCA 60  
Db 1 MetAsnTyrLysLysIlePheValSerSerAlaLeuLeuSerLeuMetSerIleLeuPro 20

Qy 61 TATCAGCTCTTTGCGAGATCTGTAGGTTCAAGA-----ACTAATGATAACAAGAA 111  
Db 21 TyrGlnSerPheAlaAspProValThrSerAsnAspThrGlyIleAsnAspSerArgGlu 40

Qy 112 GGCTTCTACATAGTGCAGATCAATCCAGTATATACACATTTAGNAAATTTCTTGCT 171  
Db 41 GlyPheTyrIleSerValLysTyrAsnProSerIleSerHisPheArgLysPheSerAla 60

Qy 172 GAAGAACTCTATTAAATGGAACAAATTCCTCACTAAAAAGTTTTCGGACTAAAGAA 231  
Db 61 GluGluAlaProIleAsnGlyAsnThrSerIleThrLysLysValPheGlyLeuLysLys 80

Qy 232 GATGGTGATATACAAAAAAAGACGATTTTACAAGAGTAGCTCCAGGCAATGATTTCAA 291  
Db 81 AspGlyAspIleAlaGlnSerAlaAsnPheAsnArgThrAspProAlaLeuGluPheGln 100

Qy 292 AATAACTTAATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGAGCGGACCAAGATA 351  
Db 101 AsnAsnLeuIleSerGlyPheSerGlySerIleGlyTyrAlaMetAspGlyProArgIle 120

Qy 352 GAATCTGAAGCTGATATCAAAATTTAATCCAAAAACACCGATTAACAGATGACTGAT 411  
Db 121 GluLeuGluAlaLysTyrGlnSerAlaAsnPheAsnArgThrAspProAlaLeuGluPheGln 140

Qy 412 AATGCTGAATATCTATAAACAATTTTGCATTTATCTGTAATTAATGATGCTCTCA 471  
Db 141 SerGlyAspTyrTyrLysTyrPheGlyLeuSerArgGluAspAlaIleAlaAspLysLys 160

Qy 472 TATGTAGTACTTAAAAATGACGATAACCTTTTATGTTCATTGTTTATGTTTACTTGT 531  
Db 161 TyrValValLeuLysAsnGluGlyIleThrPheMetSerLeuMetValAsnThrCysTyr 180

Qy 532 GACATCTGCTCAGTAACCTCTTGACGTTTGATATCTTTGGCGGAGAAATGGGAATGAG 591  
Db 181 AspIleThrAlaGluGlyValProPheIleProTyrAlaCysAlaGlyValGlyAlaAsp 200

Qy 592 CTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAATAGGTATT 651  
Db 200 81 AspGlyAspIleAlaGlnSerAlaAsnPheAsnArgThrAspProAlaLeuGluPheGln 100

Qy 651 CTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAATAGGTATT 651  
Db 201 LeuIleAsnValPheLysAspPheAsnLeuLysPheSerTyrGlnGlyLysIleGlyLe 220

Qy 652 AGTTACCTATCACACGAGAGTCTGCAATTTATGTTGATATCTACCATGCGGTATT 711  
Db 221 SerTyrProIleThrProGluValSerAlaPheIleGlyTyrTyrHisGlyValIle 240

Qy 712 GGTAATAATTTGAGAAGATACCTGTAATCTCTGTAATTAATGATGCTCTCTCA 771  
Db 241 GlyAsnAsnPheAsnLysIleProValIleThrProValIleLeuGluGlyAlaProGln 260

Qy 772 ACCATCTGCTCAGTAACCTCTTGACGTTTGATATCTTTGGCGGAGAAATGGGAATGAG 831  
Db 261 ThrThrSerAlaLeuValThrIleAspThrGlyTyrPheGlyGlyGlyLeuValGlyValArg 280

Qy 832 TTCACCTTC 840  
Db 281 PheThrPhe 283

## RESULT 10

US-09-846-808-14

; Sequence 14, Application US/09846808

Patent No. US20020064531A1  
GENERAL INFORMATION: David H.  
APPLICANT: Walker, David H.  
APPLICANT: Yu, Xu-Jie  
TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane  
TITLE OF INVENTION: Protein Multigene Family  
FILE REFERENCE: D6311  
CURRENT APPLICATION NUMBER: US/09/846,808  
PRIOR FILING DATE: 2001-05-01  
PRIOR APPLICATION NUMBER: 60/201,035  
PRIOR FILING DATE: 2000-05-01  
NUMBER OF SEQ ID NOS: 53  
SEQ ID NO 14  
LENGTH: 283  
TYPE: PRT  
ORGANISM: Ehrlichia chaffeensis  
FEATURE:  
OTHER INFORMATION: P28-14 Outer Membrane Protein of  
OTHER INFORMATION: Ehrlichia chaffeensis  
US-09-846-808-14

## Alignment Scores:

Pred. No.: 1,08e-121 Length: 283  
Score: 1202.50 Matches: 225  
Percent Similarity: 88.69% Conservative: 26  
Best Local Similarity: 79.51% Mismatches: 29  
Query Match: 81.20% Indels: 3  
DB: 9 Gaps: 1

US-10-062-624-41 (1-840) x US-09-846-808-14 (1-283)

Qy 1 ATGAATTATAAGAAAATTCAGTAAGAGCGCGTTAATCTCATTAATGTCAATCTTACCA 60  
Db 1 MetAsnTyrLysLysIlePheValSerSerAlaLeuLeuSerLeuMetSerIleLeuPro 20

Qy 61 TATCAGCTCTTTGCGAGATCTGTAGGTTCAAGA-----ACTAATGATAACAAGAA 111  
Db 21 TyrGlnSerPheAlaAspProValThrSerAsnAspThrGlyIleAsnAspSerArgGlu 40

Qy 112 GGCTTCTACATAGTGCAGATCAATCCAGTATATACACATTTAGNAAATTTCTTGCT 171  
Db 41 GlyPheTyrIleSerValLysTyrAsnProSerIleSerHisPheArgLysPheSerAla 60

Qy 172 GAAGAACTCTATTAAATGGAACAAATTCCTCACTAAAAAGTTTTCGGACTAAAGAA 231  
Db 61 GluGluAlaProIleAsnGlyAsnThrSerIleThrLysLysValPheGlyLeuLysLys 80

Qy 232 GATGGTGATATACAAAAAAAGACGATTTTACAAGAGTAGCTCCAGGCAATGATTTCAA 291  
Db 81 AspGlyAspIleAlaGlnSerAlaAsnPheAsnArgThrAspProAlaLeuGluPheGln 100

Qy 292 AATAACTTAATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGAGCGGACCAAGATA 351  
Db 101 AsnAsnLeuIleSerGlyPheSerGlySerIleGlyTyrAlaMetAspGlyProArgIle 120

Qy 352 GAATCTGAAGCTGATATCAAAATTTAATCCAAAAACACCGATTAACAGATGACTGAT 411  
Db 121 GluLeuGluAlaLysTyrGlnSerAlaAsnPheAsnArgThrAspProAlaLeuGluPheGln 140

Qy 412 AATGCTGAATATCTATAAACAATTTTGCATTTATCTGTAATTAATGATGCTCTCA 471  
Db 141 SerGlyAspTyrTyrLysTyrPheGlyLeuSerArgGluAspAlaIleAlaAspLysLys 160

Qy 472 TATGTAGTACTTAAAAATGACGATAACCTTTTATGTTCATTGTTTATGTTTACTTGT 531  
Db 161 TyrValValLeuLysAsnGluGlyIleThrPheMetSerLeuMetValAsnThrCysTyr 180

Qy 532 GACATCTGCTCAGTAACCTCTTGACGTTTGATATCTTTGGCGGAGAAATGGGAATGAG 591  
Db 181 AspIleThrAlaGluGlyValProPheIleProTyrAlaCysAlaGlyValGlyAlaAsp 200

Qy 592 CTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAATAGGTATT 651  
Db 200 81 AspGlyAspIleAlaGlnSerAlaAsnPheAsnArgThrAspProAlaLeuGluPheGln 100

Db 201 LeuileAenValPheLysAspPheAenLeuLysPheSerTyrGlnGlyLysleGlyIle 220  
Qy 652 AGTTACCTTATCACACAGAGTCTCTGCATTATTGTGTGATACACTACCTGGCGTTATT 711  
Db 221 SerTyrProIleThrProGluValSerAlaPheIleGlyGlyTyrTyrHisGlyValIle 240  
Qy 712 GGTAAATAATTGAGAAGATACCTGTAACTCTCTGTAGTATTAAATGATGCTCCCTCAA 771  
Db 241 GlyAenAenPheAenLysleProValIleThrProValValLeuGluGlyAlaProGln 260  
Qy 772 ACCACATCTGCTTACAGTAACCTCTGACCTTGATACCTTTGGCGGAGAAATGGAATGAGG 831  
Db 261 ThrThrSerAlaLeuValThrIleAspThrGlyTyrPheGlyGlyGluValGlyValArg 280  
Qy 832 TTCACCTTC 840  
Db 281 PheThrPhe 283

RESULT 11  
US-10-062-624-10  
; Sequence 10, Application US/10062624  
; Patent No. US20020115840A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-Kilodalton Immunodominant Protein  
; FILE REFERENCE: Genes of Ehrlichia canis and Uses Thereof  
; CURRENT APPLICATION NUMBER: US/10/062,624  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: 09/660,587  
; PRIOR FILING DATE: 2000-09-12  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 10  
; LENGTH: 283  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1B  
US-10-062-624-10

Alignment Scores:  
Pred. No.: 1,08e-121 Length: 283  
Score: 1202.50 Matches: 225  
Percent Similarity: 88.69% Conservative: 26  
Best Local Similarity: 79.51% Mismatches: 29  
Query Match: 81.20% Indels: 3  
DB: 12 Gaps: 1

US-10-062-624-41 (1-840) x US-10-062-624-10 (1-283)  
Qy 1 ATGAATTATAAGAAATTTCTAGTAAGAGCGGTTAATCTCATTAATGTCATCTTACCA 60  
Db 1 MetAenTyrLyslePheValSerAlaLeuLeuSerLeuLeuSerLeuLeuPro 20  
Qy 61 TATGAGTCTTTGCGAGATCTGTAGTTCAGA-----ACTAATGATAACAAGAA 111  
Db 21 TyrGlnSerPheAlaAspProValThrSerAenAspThrGlyIleAenAspSerArgGlu 40  
Qy 112 GCCTTCTACATTAGTGCAAGTACCAATCAAGTATATCACACTTTAGAAAATCTCTGCT 171  
Db 41 GlyPheTyrIleSerValLysTyrAenProSerIleSerHisPheArgLysPheSerAla 60  
Qy 172 GAAGAACTCTTATTAAATGGACAAATTTCTCTCACTAAATAAGTTTCGGACTAAAGAAA 231  
Db 61 GluGluAlaProIleAenGlyAenThrSerIleThrLysLysValPheGlyLeuLysLys 80  
Qy 232 GATGGTGATATACAAAAAGACATTTTACAGAGTAGTCCAGGATTCATTTTCAA 291  
Db 81 AspGlyAspIleAlaGlnSerAlaAenArgThrAspProAlaLeuGluPheGln 100  
Qy 292 AATAACTTAATATCAGGATTTTCAGGAAGTATTGTGTTACTTATGTCGACGACCAAGATA 351

Db 101 AsnAenLeuLeuSerGlyPheSerGlySerIleGlyTyrAlaMetAspGlyProArgIle 120  
Qy 352 GAATCGAGCTGCATATCAACAAATTTAATCCAAAAACACCGATAACAATGATGATCAT 411  
Db 121 GluLeuGluAlaLysPheAenLysPheAenProAspAenAsnAspThrAsn 140  
Qy 412 AATGCTGAATCTATAACATTTTGCATTATCTCTGTAAGATGCAATGGAAGATCAGCAA 471  
Db 141 SerGlyAspTyrTyrLysTyrPheGlyLeuSerArgGluAspAlaIleAlaAspLysLys 160  
Qy 472 TATCTAGTACTTAAAAATGACGGCATAAATTTTATGTTCATTGATGGTTAATGATCTAT 531  
Db 161 TyrValValLeuLysAenGluGlyIleThrPheMetSerLeuMetValAenThrCysTyr 180  
Qy 532 GACATTACAGCTGAAGAGATCTTTCTGCATATATGATATGATGATGATGATGATGATGAT 591  
Db 181 AspIleThrAlaGluGlyValProPheIleProTyrAlaCysAlaGlyValGlyAlaAsp 200  
Qy 592 CTTATCACTATTTTAAAGACCTCAATCAAAATTTGCTTACCAGGAAGAAATAGGTATT 651  
Db 201 LeuIleAenValPheLysAspPheAenLeuLysPheSerTyrGlnGlyLysleGlyIle 220  
Qy 652 AGTTACCTTATCACACAGAGTCTCTGCATTTTATGTTGATGATGATGATGATGATGATGAT 711  
Db 221 SerTyrProIleThrProGluValSerAlaPheIleGlyGlyTyrTyrHisGlyValIle 240  
Qy 712 GGTAAATAATTGAGAAGATACCTGTAACTAATCTCTGTAGTATTAAATGATGATGATGAT 771  
Db 241 GlyAenAenPheAenLysleProValIleThrProValValLeuGluGlyAlaProGln 260  
Qy 772 ACCACATCTGCTTCAAGTAACTCTTGCATGTTGATGATGATGATGATGATGATGATGAT 831  
Db 261 ThrThrSerAlaLeuValThrIleAspThrGlyTyrPheGlyGlyGluValGlyValArg 280  
Qy 832 TTCACCTTC 840  
Db 281 PheThrPhe 283

RESULT 12  
US-10-059-964-4  
; Sequence 4, Application US/10059964  
; Patent No. US20020120115A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohashi, No. US20020120115A1  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; TITLE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/059,964  
; CURRENT FILING DATE: 2002-01-28  
; EARLIER APPLICATION NUMBER: 09/314,701  
; EARLIER FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 283  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
US-10-059-964-4

Alignment Scores:  
Pred. No.: 1,08e-121 Length: 283  
Score: 1202.50 Matches: 225  
Percent Similarity: 88.69% Conservative: 26  
Best Local Similarity: 79.51% Mismatches: 29  
Query Match: 81.20% Indels: 3  
DB: 12 Gaps: 1

US-10-062-624-41 (1-840) x US-10-059-964-4 (1-283)  
Qy 1 ATGAATTATAAGAAATTTCTAGTAAGAGCGGTTAATCTCATTAATGTCATCTTACCA 60



1 MetAsnTyrLysLysIlePheValSerSerAlaLeulleSerLeuMetSerIleLeuPro 20  
61 TATCAGTCTTTTGCAGATCCTGTAGTTCAAGA-----ACTAATCATACAAAGAA 111  
21 TyrGlnSerPheAlaAspProValThrSerAsnAspThrGlyIleAsnAspSerArgGlu 40  
112 GGCCTCTACATTAGTGCAGAGTACAAATCCAAAGTATATCACACTTTTAGAAAAATCTCTGCT 171  
41 GlyPheTyrIleSerValLysTyrAsnProSerIleSerHisPheArgLysPheSerAla 60  
172 GAAGAAACTCTCTATTAAATGGAGCAAAATCTCTCACTAAAAAAGTTTTCGGACTAAAGAA 231  
61 GluGluAlaProIleAsnGlnYAsnThrSerIleThrLysLysValPheGlyLeuLysLys 80  
232 GATGGTGATATAACAAAAAAAGACGATTTTCAAGAGTAGTCTCCAGGCATTCGATTTTCAA 291  
81 AspGlyAspIleAlaGlnSerAlaAsnPheAsnArgThrAspProAlaLeuGluPheGln 100  
292 AATAACTTAATATCAGGATTTTCAGGAAGTATGGTTACTCTATCGACGGACCAAGAAATA 351  
101 AsnAsnLeulleSerGlyPheSerGlySerIleGlyTyrAlaMetAspGlyProArgIle 120  
352 GAACCTTGAAGCTGCATATCAACAATTTAATCCAAAAAACCCGATAACAATCATACTGAT 411  
121 GluLeuGluAlaIaIatyrGlnLysPheAspAlaLysAsnProAspAsnAsnAspThrAsn 140  
412 AATGGTGAATACTATAACATTTTCCATATCTCTGTAAGATGCAATGGAAGATCAGCAAA 471  
141 SerGlyAspTyrTyrLysTyrPheGlyLeuSerArgGluAspAlaIleAlaAspLysLys 160  
472 TATGTAGTACTTAAAAATGACGGCAATACTTTTATGTCTCATTCGATGTTAATACTTGCTAT 531  
161 TyrValValLeuLysAsnGluGlyIleThrPheMetSerLeuMetValAsnThrCysTyr 180  
532 GACATTACAGCTGAAGGAGTATCTTTTCGTACCATATGCATGTGCAGGTATAGGACGAGAT 591  
181 AspIleThrAlaGluGlyValProPheIleProTyrAlaCysAlaGlyValGlyAlaAsp 200  
592 CTTATCACTATTTTAAAGACCTCAATCTAAATTTTCTTACCACAGGAAAAATAGGTATT 651  
201 LeulleAsnValPheLysAspPheAsnLeuLysPheSerTyrGlnGlyLysIleGlyIle 220  
652 AGTTACCCCTATCACACCAAGAGTCTCTGCATTTATTGGTGGATACTACCATGGCGTTATT 711  
221 SerTyrProIleThrProGluValSerAlaPheIleGlyTyrTyrHisGlyValIle 240  
712 GGTAAATAATTTGAGAAGATACCTGTATAACTCTCTGTAGTATTTAAATGATGCTCTCTCAA 771  
241 GlyAsnAsnPheAsnLysIleProValIleThrProValValLeuGluGlyAlaProGln 260  
772 ACCACATCTGCTTCAGTAACTCTTCACGCTGGATACTTTGGCGGAGAAATTCGAATGAGG 831  
261 ThrThrSerAlaLeuValThrIleAspThrGlyTyrPheGlyGlyGluValGlyValArg 280  
832 TTCACCTTC 840  
281 PheThrPhe 283

RECIT. 13

RESULT 13  
 US-10-062-051-9  
 ; Sequence 9, Application US/10062051  
 ; Publication No. US20030073095A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Walker, David H.  
 ; APPLICANT: McBride, Jere W.  
 ; APPLICANT: Yu, Xue-Jie  
 ; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
 ; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof  
 ; FILE REFERENCE: D6152CIP2  
 ; CURRENT APPLICATION NUMBER: US/10/062,051  
 ; CURRENT FILING DATE: 2002-01-31  
 ; PRIOR APPLICATION NUMBER: US/09/660,587  
 ; PRIOR FILING DATE: 2000-09-12

```

, PRIOR APPLICATION NUMBER: 09/261,358
,
, PRIOR FILING DATE: 1999-03-03
,
, NUMBER OF SEQ ID NOS: 46
,
, SEQ ID NO 9
,
, LENGTH: 281
,
, TYPE: PRT
,
, ORGANISM: Ehrlichia chaffeensis
,
, FEATURE:
,
, OTHER INFORMATION: amino acid sequence of E. chaffeensis p28

```

Alignment Scores:		
Pred. No.:	3,456-61	Length:
Score:	644.50	Matches:
Percent Similarity:	62.10%	Conservative:
Best Local Similarity:	48.11%	Mismatches:
Query Match:	43.52%	Indels:
DB:	9	Gaps:
		281
		140
		41
		89
		21
		6

US-10-062-624-41 (1-840) x US-10-062-051-9 (1-281)

Qy	1	ATGAATTATAAGAAAAATCTTAGTAAGAAGCCGGTTAAATCTCATTTAATGTCAAATCTTATACCA	60
Db	1	MetAsnTyrLysValPheIleThrSerAlaLeuIleSerLeuIleSerLeuPro	20
Qy	61	TATCAGTCTTTTCAGATCCTGTTAGGTTCAAGAAGTAATGATAACAAGAAGCGCTTCTTAC	120
Db	21	GlyValSerPheSerAspProAlaGlySerGlyIleAsnGlyAsn-----PheTyr	37
Qy	121	ATTAGTGCAAAGTACAATCCAAAGTATATCACACTTTTACAAAATTCCTGCTCAAGAAACT	180
Db	38	IleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu---	56
Qy	181	CCTATTAAATGGAACAAAATTCCTCCTCACTAAAAAGTTTTCGGACTAAGAAGAAGATGGTGTAT	240
Db	57	-----GluArgAsnThrThrValGlyValPheGlyLeuLysGlnAsnTrpAsp	72
Qy	241	-----ATACAAAAAAGACGATTTTACAGAGTAGCTCCAGGCAT	282
Db	73	GlySerAlaIleSerAsnSerSerProAsnAspValPheThrValSerAsnTyrSerPhe	92
Qy	283	GATTTTCAAATAACTTAATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGACACGA	342
Db	93	LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetAspGly	112
Qy	343	CCAAGATAGAACTTGAGCTGCATATCAACAATTTAATCAAAAACACCGATTAACAAT	402
Db	113	ProArgIleGluLeuGluIleValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn	132
Qy	403	GATACTGATAATGGTGAATACATAAACAATTTTCGATTATCTCGTAAAGATGCAATGGAA	462
Db	133	-----TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerHisAsnSerAlaAlaAsp	150
Qy	463	-----GATCAGCAATGTAGTACTTAAAAATGACGGCATTAACCTTTTATGTCA	510
Db	151	MetSerSerAlaSerAsnAsnPheValPheLeuLysAsnGluGlyLeuLeuAspIleSer	170
Qy	511	TTGATGTTTAATCTTGCTATGACATTAACAGCTGAAGAGATATCTTTCGTACCATATGCA	570
Db	171	PheMetLeuAsnAlaCysTyrAspValValGlyGluGlyIleProPheSerProTyrIle	190
Qy	571	TGTCACGGTATAGCAGCAGATCTTATCACTATTTTAAAGACCTCAACTAAAATTTGCT	630
Db	191	CysAlaGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSer	210
Qy	631	TACCAAGGAAAAATAGGTATTAGTTACCTATCACACAGCAAGTCTCTGCATTTATTGCT	690
Db	211	TyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSerValPheIleGly	230
Qy	691	GGATACTACCATGGCGTTATTGGTAAATAAATTGAGAAGATACCTGTATAACTCCTGTA	750
Db	231	GlyHisPheHisValIleGlyAsnGluPheArgAspIleProThrIleIleProThr	250

QY 751 GTATTAAATGATGCTCTCAA---ACCACATCTGCTTCAAGTAACCTCTTGAAGTGGATAC 807  
Db 251 GlySerThrLeuAlaGlyLysGlyAsnTyrProAlaIleValIleLeuAspValCysHis 270  
QY 808 TTTCGCGAGAAATTGGAATGAGTTCACCTTC 840  
Db 271 PheGlyIleGluLeuGlyGlyArgPheAlaPhe 281

RESULT 14  
US-10-284-986-19  
; Sequence 19, Application US/10284986  
; Publication No. US20030091588A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: Yu, Xu-Jie  
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane  
; FILE REFERENCE: D6311D1  
; CURRENT APPLICATION NUMBER: US/10/284,986  
; CURRENT FILING DATE: 2002-10-31  
; PRIOR APPLICATION NUMBER: 09/846,808  
; PRIOR FILING DATE: 2001-05-01  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 19  
; LENGTH: 281  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: P28-19 Outer Membrane Protein of  
; OTHER INFORMATION: Ehrlichia chaffeensis  
US-10-284-986-19

Alignment Scores:  
Pred. No.: 3,45e-61 Length: 281  
Score: 644.50 Matches: 140  
Percent Similarity: 62.20% Conservative: 41  
Best Local Similarity: 48.11% Mismatches: 89  
Query Match: 43.52% Indels: 21  
Gaps: 6

US-10-062-624-41 (1-840) x US-10-284-986-19 (1-281)

QY 1 ATGAATTATAAGAAATTTCTAGTAAGACGCGTAAATCTCAATTAATGTCATCTTACCA 60  
Db 1 MetAsnTyrLysValPheIleThrSerAlaLeuIleSerLeuIleSerLeuPro 20  
QY 61 TATCAGTCTTTGACAGCTCTGAGTTCAGAACTAATGATAACAAAGAGGCTTCTAC 120  
Db 21 GlyValSerPheSerAspProAlaGlySerGlyIleAsnGlyAsn-----PheTyr 37  
QY 121 ATTAGTCAAGTACAATCCAAGTATATACACTTTAGAAATTTCTCTGCTGAAGAACT 180  
Db 38 IleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu--- 56  
QY 181 CCTATTATGGAACAATTTCTCTCAATAAAGTTTCGGAATAAGTAAGAGATGGTAT 240  
Db 57 -----GluArgAsnThrThrValGlyValPheGlyLeuLysIleAsnTrpAsp 72  
QY 241 -----ATACAAAAAAGACGATTTTACAAAGTAGTAGCTCCAGGCATT 282  
Db 73 GlySerAlaIleSerAsnSerSerProAsnAspValPheThrValSerAsnTyrSerPhe 92  
QY 283 GATTTTCAAAATAACTTAATATATCAGATTTTTCAGGAAGTATTGGTTACTCTATGACCGA 342  
Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetAspGly 112  
QY 343 CCAAGATAGACTTGAAGCTCATATCAACAATTTAATCCAAAAACACCGATACAAT 402  
Db 113 ProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn 132  
QY 403 GATACTGATAATGGTGAATCTATAACATTTTGCATTATCTCGTAAAGATGCAATGGAA 462  
Db 133 -----TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerHisAsnSerAlaAlaAsp 150

QY 463 -----GATCAGCAATATGTAGTACTTAAATGACGCGATAAATCTTTTATGTCA 510  
Db 151 MetSerSerLalaSerAsnAsnPheValPheLeuLysAsnGluGlyLeuLeuAspIleSer 170  
QY 511 TTGATGTTAATACTTCTGATGACATTACAGCTGAAGGAGTATCTTTCGTACCATATGCA 570  
Db 171 PheMetLeuAsnAlaCysTyrAspValValGlyGluGlyIleProPheSerProTyrIle 190  
QY 571 TGTCCAGGTATPAGGAGAGATCTTATCACTATTTTAAAGACCTCAATCTTAAATTTGCT 630  
Db 191 CysAlaGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSer 210  
QY 631 TACCAAGAAAAATAGTATTTACCTTATACCTATCACACAGAGTCTCTGCATTTTGGT 690  
Db 211 TyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSerValPheIleGly 230  
QY 691 GGATACTACCATGCGCTTATTGGTAATAAATTTGAGAAGATACCTGTAATACCTCTGTA 750  
Db 231 GlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrIleProThr 250  
QY 751 GTATTAAATGATGCTCTCAA---ACCACATCTGCTTCAAGTAACCTCTTGAAGTGGATAC 807  
Db 251 GlySerThrLeuAlaGlyLysGlyAsnTyrProAlaIleValIleLeuAspValCysHis 270

QY 808 TTTCGCGAGAAATTGGAATGAGTTCACCTTC 840  
Db 271 PheGlyIleGluLeuGlyGlyArgPheAlaPhe 281

RESULT 15  
US-10-062-920-9  
; Sequence 9, Application US/10062920  
; Publication No. US20030096250A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: US/10/062,920  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: US/09/660,587  
; PRIOR FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 05/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 9  
; LENGTH: 281  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of E. chaffeensis P28  
US-10-062-920-9

Alignment Scores:  
Pred. No.: 3,45e-61 Length: 281  
Score: 644.50 Matches: 140  
Percent Similarity: 62.20% Conservative: 41  
Best Local Similarity: 48.11% Mismatches: 89  
Query Match: 43.52% Indels: 21  
Gaps: 6

US-10-062-624-41 (1-840) x US-10-062-920-9 (1-281)

QY 1 ATGAATTATAAGAAATTTCTAGTAAGACGCGTAAATCTCAATTAATGTCATCTTACCA 60  
Db 1 MetAsnTyrLysValPheIleThrSerAlaLeuIleSerLeuIleSerLeuPro 20  
QY 61 TATCAGTCTTTGACAGTCTGAGTTCAGAACTAATGATAACAAAGAGGCTTCTAC 120  
Db 21 GlyValSerPheSerAspProAlaGlySerGlyIleAsnGlyAsn-----PheTyr 37



## RESULT 17

US-10-062-624-9  
; Sequence 9, Application US/10062624  
; Patent No. US20020115840A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-Kilodalton Immunodominant Protein  
; FILE REFERENCE: Genes of Ehrlichia canis and Uses Thereof  
; FILE REFERENCE: D6152CIP2/D1  
; CURRENT APPLICATION NUMBER: US/10/062,624  
; PRIORITY FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: 09/660,587  
; PRIOR FILING DATE: 2000-09-12  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 9  
; LENGTH: 281  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of E. chaffeensis P28  
US-10-062-624-9

Alignment Scores:  
Pred. No.: 3,456-61 Length: 281  
Score: 644.50 Matches: 140  
Percent Similarity: 62.20% Conservative: 41  
Best Local Similarity: 48.11% Mismatches: 89  
Query Match: 43.52% Indels: 21  
DB: 12 Gaps: 6

US-10-062-624-41 (1-840) x US-10-062-624-9 (1-281)

QY 1 ATGAATTATAAGAAAATCTAGTAAGACGCGTTAATCTCATTAATGTCTTAATCTTACCA 60  
DB 1 MetAsnTyrLysValPheIleThrSerAlaLeuIleSerLeuLeuPro 20  
QY 61 TATCAGTCTTTGTCAGATCCTGTAGTTCACAGAACTAATGATAACAAGAGGCTTCTAC 120  
DB 21 GlyValSerPheSerAspProAlaGlySerGlyIleAsnGlyAsn-----PheTyr 37  
QY 121 ATTAGTCAAAAGTACAATCCAGTATATACACATTTTAGAAAATCTCTGCTGAAGAACT 180  
DB 38 IleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu--- 56  
QY 181 CCTATTATGGAACAAATTTCTCTCACTAAAGATTTTCGGACTAAAGAAAGATGGTAT 240  
DB 57 -----GluArgAsnThrThrValGlyValPheGlyLeuLysGlnAsnTrpAsp 72  
QY 241 -----ATAACAAAAAGACGATTTTACAGAGTAGTCTCCAGGCATT 282  
DB 73 GlySerAlaIleSerAsnSerProAsnAspValPheThrValSerAsnTyrSerPhe 92  
QY 283 GATTTCACAAATAACTTAATACAGATTTTCAGGAGATGATTGGTTACTCTATGACCGGA 342  
DB 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetAspGly 112  
QY 343 CCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACACCCGATACAAT 402  
DB 113 ProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn 132  
QY 403 GATACCTGATATGGTGAATACATATAACATTTTGGCATTTATCTCGTAAAGATGCAATGAA 462  
DB 133 -----TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerHisAsnSerAlaAlaAsp 150  
QY 463 -----GATCAGCAATATCTAGTACTTAAATAATGACGGCATACTTTTATGCA 510  
DB 151 MetSerSerAlaSerAsnAsnValPheLeuLysAsnGluGlyLeuLeuAspIleSer 170  
QY 511 TTGATGGTTAACTTACATGATACATGATGAGGAGTATCTTCGTACCATATGCA 570  
DB 171 PheMetLeuAsnAlaCysTyrAspValValGlyGlyGluGlyIleProPheSerProTyrIle 190

QY 571 TGTCCAGGTATAGCAGACAGCTTATCAGTATTTTAAAGACCTCAATCTAAATTTGCT 630  
DB 191 CysAlaGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSer 210  
QY 631 TACCAAGGAAAAATAGGTATTAGTTACCTATCACACCAGAGTCTCTGCAATTTATGGT 690  
DB 211 TyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSerValPheIleGly 230  
QY 691 GGATACCTACCATGCGGTTATTGTTAAATAATTGAGAAGATACCTGTAAATACCTCTGTA 750  
DB 231 GlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrIleIleProThr 250  
QY 751 GTATTAATGATGCTCTCTCAA---ACCACATCTCTTCAGTAACTCTTGAGTTGGATAC 807  
DB 251 GlySerThrLeuAlaGlyLysGlyAsnTyrProAlaIleValIleLeuAspValCysHis 270  
QY 808 TTTCGCGGAGAAATTTGGAATGAGGTTTCACCTTC 840  
DB 271 PheGlyIleGluLeuGlyGlyArgPheAlaPhe 281

## RESULT 18

US-10-314-639-2  
; Sequence 2, Application US/10314639  
; Publication No. US20030103991A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20030103991A1  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; TITLE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/314,639  
; CURRENT FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US/09/314,701  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 281  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
US-10-314-639-2

Alignment Scores:  
Pred. No.: 5,686-61 Length: 281  
Score: 642.50 Matches: 139  
Percent Similarity: 62.20% Conservative: 42  
Best Local Similarity: 47.77% Mismatches: 89  
Query Match: 43.38% Indels: 21  
DB: 9 Gaps: 6

US-10-062-624-41 (1-840) x US-10-314-639-2 (1-281)

QY 1 ATGAATTATAAGAAAATTTAGTAAGACGCGTTAATCTCATTAATGTCTTAATCTTACCA 60  
DB 1 MetAsnTyrLysLysValPheIleThrSerAlaLeuIleSerLeuLeuPro 20  
QY 61 TATCAGTCTTTGTCAGATCCTGTAGTTCACAGAACTAATGATAACAAGAGGCTTCTAC 120  
DB 21 GlyValSerPheSerAspProAlaGlySerGlyIleAsnGlyAsn-----PheTyr 37  
QY 121 ATTAGTCAAAAGTACAATCCAGTATATACACATTTTAGAAAATCTCTGCTGAAGAACT 180  
DB 38 IleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu--- 56  
QY 181 CCTATTATGGAACAAATTTCTCTCACTAAAGATTTTCGGACTAAAGAAAGATGGTAT 240  
DB 57 -----GluArgAsnThrThrValGlyValPheGlyLeuLysGlnAsnTrpAsp 72  
QY 241 -----ATAACAAAAAGACGATTTTACAGAGTAGTCTCCAGGCATT 282  
DB 73 GlySerAlaIleSerAsnSerProAsnAspValPheThrValSerAsnTyrSerPhe 92



```
; APPLICANT: Alleman, Arthur R.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and
; FILE REFERENCE: Methods of use
; CURRENT APPLICATION NUMBER: US/10/062,994
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 09/533,662
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 09/337,827
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 08/953,326
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 08/733,230
; PRIOR FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-10-062-994-18

Alignment Scores:
Pred. No.: 1,45e-59 Length: 276
Score: 629.50 Matches: 137
Percent Similarity: 62.24% Conservative: 41
Best Local Similarity: 47.90% Mismatches: 87
Query Match: 42.51% Indels: 21
DB: 12 Gaps: 6

US-10-062-624-41 (1-840) x US-10-062-994-18 (1-276)

Qy 1 ATGAATATAGAAAATCTAGTAGAAGCGGTTAATCTCAATTAATCTCAATCTTACCA 60
Db 1 MetAsnTyrLysLeuValPheThrSerAlaLeuLeuSerLeuLeuSerLeuPro 20
Qy 61 TATCAGTCTTTGCGAGATCTGTAGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120
Db 21 GlyValSerPheSerAspProAlaGlySerGlyLeuAsnGlyAsn-----PheTyr 37
Qy 121 ATTAGTCAAAAGTACAATCTCAAGTATATACACATTTTAGAAAATCTCTGCTGAAGAACT 180
Db 38 IleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu--- 56
Qy 181 CCTATTATGAACAATAATCTCTCAATAAAGTTTCGAGCTAAAGAAAGATGGTGAAT 240
Db 57 -----GluArgAsnThrThrValGlyValPheGlyLeuLysGlnAsnTrpAsp 72
Qy 241 -----ATACAAAAAGAGGATTTTACAAGAGTAGTCCAGGCATT 282
Db 73 GlySerAlaIleSerAsnSerSerProAsnAspValPheThrValSerAsnTyrSerPhe 92
Qy 283 GATTTTCAAAATTAATATATCAGGATTTTTCAGGAAGTATTGGTTACTTATGACGGA 342
Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetAspGly 112
Qy 343 CAAAGATAGAACTTGAAGCTGCATATCAACAATTAATCAAAAACACCGATACAAT 402
Db 113 ProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn 132
Qy 403 GATACGTGAATGGTGAATACATATAAATTTTGCATTATCTCGTAAAGATGCAATGAA 462
Db 133 -----TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerHisAsnSerAlaAlaAsp 150
Qy 463 -----GATCAGCAATATATAGTACTTAAATAATGACGGCATATCTTTTGTCA 510
Db 151 MetSerSerAlaSerAsnAsnPheValPheLeuLysAsnGluGlyLeuLeuAspIleSer 170
Qy 511 TTGATGGTTAATACTTGTATGACATATACAGCTGAAGGAGTATCTTTCGTACCATATCA 570
Db 171 PheMetLeuAsnAlaCysTyrAspValValGlyGluGlyIleProPheSerProTyrIle 190
Qy 571 TGTGCGAGTATAGGACGAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCT 630
Db 21 GlyValSerPheSerAspProAlaGlySerGlyLeuAsnGlyAsn-----PheTyr 37

; APPLICANT: Alleman, Arthur R.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and
; FILE REFERENCE: Methods of use
; CURRENT APPLICATION NUMBER: US/10/062,994
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 09/533,662
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 09/337,827
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 08/953,326
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 08/733,230
; PRIOR FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-10-062-994-18

Alignment Scores:
Pred. No.: 1,45e-59 Length: 276
Score: 629.50 Matches: 137
Percent Similarity: 62.24% Conservative: 41
Best Local Similarity: 47.90% Mismatches: 87
Query Match: 42.51% Indels: 21
DB: 12 Gaps: 6

US-10-062-624-41 (1-840) x US-10-062-994-18 (1-276)

Qy 1 ATGAATATAGAAAATCTAGTAGAAGCGGTTAATCTCAATTAATCTCAATCTTACCA 60
Db 1 MetAsnTyrLysLeuValPheThrSerAlaLeuLeuSerLeuLeuSerLeuPro 20
Qy 61 TATCAGTCTTTGCGAGATCTGTAGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120
Db 21 GlyValSerPheSerAspProAlaGlySerGlyLeuAsnGlyAsn-----PheTyr 37
Qy 121 ATTAGTCAAAAGTACAATCTCAAGTATATACACATTTTAGAAAATCTCTGCTGAAGAACT 180
Db 38 IleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu--- 56
Qy 181 CCTATTATGAACAATAATCTCTCAATAAAGTTTCGAGCTAAAGAAAGATGGTGAAT 240
Db 57 -----GluArgAsnThrThrValGlyValPheGlyLeuLysGlnAsnTrpAsp 72
Qy 241 -----ATACAAAAAGAGGATTTTACAAGAGTAGTCCAGGCATT 282
Db 73 GlySerAlaIleSerAsnSerSerProAsnAspValPheThrValSerAsnTyrSerPhe 92
Qy 283 GATTTTCAAAATTAATATATCAGGATTTTTCAGGAAGTATTGGTTACTTATGACGGA 342
Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetAspGly 112
Qy 343 CAAAGATAGAACTTGAAGCTGCATATCAACAATTAATCAAAAACACCGATACAAT 402
Db 113 ProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn 132
Qy 403 GATACGTGAATGGTGAATACATATAAATTTTGCATTATCTCGTAAAGATGCAATGAA 462
Db 133 -----TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerHisAsnSerAlaAlaAsp 150
Qy 463 -----GATCAGCAATATATAGTACTTAAATAATGACGGCATATCTTTTGTCA 510
Db 151 MetSerSerAlaSerAsnAsnPheValPheLeuLysAsnGluGlyLeuLeuAspIleSer 170
Qy 511 TTGATGGTTAATACTTGTATGACATATACAGCTGAAGGAGTATCTTTCGTACCATATCA 570
Db 171 PheMetLeuAsnAlaCysTyrAspValValGlyGluGlyIleProPheSerProTyrIle 190
Qy 571 TGTGCGAGTATAGGACGAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCT 630
Db 21 GlyValSerPheSerAspProAlaGlySerGlyLeuAsnGlyAsn-----PheTyr 37
```

Qy 121 ATTAGTCAAGTACAACTCCAGTATATATACACATTTTAGAAAATTTCTCTGCTGAAGAACT 180  
Db 38 IISerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu--- 56  
Qy 181 CCTATTAAATGAACAAATTTCTCTCACTAAAAGTTTTCGAGCTAAAGAAAGATGGTAT 240  
Db 57 -----GluArgAsnThrValGlyValPheGlyLeuLysGlnAsnTrpAsp 72  
Qy 241 -----ATAACAAAAGACAGATTTTACAGAGTAGTCTCCAGGCATT 282  
Db 73 GlySerAlaIleSerAsnSerProAsnAspValPheThrValSerAsnTyrSerPhe 92  
Qy 283 GATTTTCAAAATACTTAATATCAGGATTTTCAGAGATATTGTTACTCTATGAGCGGA 342  
Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetAspGly 112  
Qy 343 CCAAGATAGAACTTGAAGCTGCATATCAACAAATTTTAAATCCAAAAACACCGATAACAAT 402  
Db 113 ProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsn 132  
Qy 403 GATACTGATAATGGTGAATACTATATAACATTTTGCATTATCTCGTAAAGATGCAATGGA 462  
Db 133 -----TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerHisAsnSerAlaAsp 150  
Qy 463 -----GATCAGCAATATGTAGTACTTAAATGACGGCATAACTTTTATGTCA 510  
Db 151 MetSerSerAlaSerAsnAsnPheValPheLeuLysAsnGluGlyLeuLeuAspIleSer 170  
Qy 511 TTGATGTTAATACTCTGATGACATTACAGCTCAAGGAGTATCTTTCGTACCATATGCA 570  
Db 171 PheMetLeuAsnAlaCysTyrAspValValGlyGluGlyIleProPheSerProTyrIle 190  
Qy 571 TGTGAGGTATGAGGAGCATCTTATCACTATTTTAAAGACCTCAATCTAAAATTTGCT 630  
Db 191 CysAlaGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSer 210  
Qy 631 TACCAAGAAAATAGTATTAGTTACCTATACACAGAGTCTCTGCAATTTATGGT 690  
Db 211 TyrGlnGlyLysLeuGlyLysSerTyrSerIleSerProGluAlaSerValPheIleGly 230  
Qy 691 GGATACCTACCATGCGTTATTGGTAATAATTTGAGAAGATACCTGTAATACTCTCTGTA 750  
Db 231 GlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrIleProThr 250  
Qy 751 GTATTAAATGATGCTCTCTCAA---ACCACATCTCTTCACTAATCTTTCAGTGTGATAC 807  
Db 251 GlySerThrLeuAlaGlyLysGlyAsnTyrProAlaIleValIleLeuAspValCysHis 270  
Qy 808 TTGGCGGAGAAATGGA 825  
Db 271 PheGlyIleGluMetGly 276

## RESULT 22

US-10-314-639-32  
; Sequence 32, Application US/10314639  
; Publication No. US20030103991A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20030103991A1  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; TITLE OF INVENTION: Chafreensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/314,639  
; PRIOR FILING DATE: 2002-12-09  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 32  
; LENGTH: 288  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
US-10-314-639-32

Alignment Scores:  
Pred. No.: 1,23e-58 Length: 288  
Score: 621.00 Matches: 135  
Percent Similarity: 62.08% Conservative: 50  
Best Local Similarity: 45.30% Mismatches: 85  
Query Match: 41.93% Indels: 28  
DB: Gaps: 7

US-10-062-624-41 (1-840) x US-10-314-639-32 (1-288)

Qy 1 ATGAATTTATAGAAAATTTCTAGTAAGAGCGGTTAATCTCATTAATGTCATCTTACCA 60  
Db 1 MetAsnCysLysArgPhePheIleAlaSerAlaLeuIleSerLeuMetSerPheLeuPro 20  
Qy 61 TATCAGTCTTTTTCAGATCTCTGAGGTTCAGAACTAATGATACAAAGAGGC---TTC 117  
Db 21 SerValSerPheSerGluSerIle-----HisGluAspAsnIleAsnGlyAsnPhe 37  
Qy 118 TACATTAGTCAAAAGTACAAATCCAAAGTATATACACATTTAGAAAATTTCTCTGCTGAAGAA 177  
Db 38 TyIleSerAlaLysTyrMetProSerAlaSerHisPheGlyValPheSerValLysGlu 57  
Qy 178 ACTCTTATTAATGGAACAAATTTCTCTCACTAAAAGTTTTCGAGCTAAAGAAAGATGGT 237  
Db 58 -----GluLysAsnThrThrThrGlyValPheGlyLeuLysGlnAspTrp 72  
Qy 238 GAT---ATAACAAAAGAGCAT-----TTC 261  
Db 73 AspGlyAlaThrIleLysAspAlaSerSerSerHisThrIleAspProSerThrIlePhe 92  
Qy 262 ACAAGAGTAGTCTCCAGCATTTGATTTTCAAAATAAATTAATATATCAGGATTTTCAGGAAGT 321  
Db 93 SerIleSerAsnTyrSerPheLysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAla 112  
Qy 322 ATTGGTTACTTATGAGCGGACCAAGATAAGACTTGAAGCTGCATATCAACAATTTAAT 381  
Db 113 IleGlyTyrSerMetGlyGlyProArgValGluPheGlyValSerTyrGluIlePheAsp 132  
Qy 382 CCAAAAACCCGATACATGATGATAAATGGTGAATGCTTATAAACAATTTTGCATTA 441  
Db 133 ValLysAsnGlnGlyAsnSerTyrLysAsnAspAla-----HisLysTyrCysAlaLeu 150  
Qy 442 TCTCGT-----AAAGATGCAATGGAAGATCAGCAATATAGTAGTACTTAAA 486  
Db 151 SerArgHisThrGlyGlyMetProGlnAlaGlyHisGlnAsnLysPheValPheLeuLys 170  
Qy 487 AATGACGCGATAACTTTTATGTCATTTGATGTTTAACTACTTGTGATGATGATGATGATG 546  
Db 171 AsnGluGlyLeuLeuAspIleSerLeuMetIleAsnAlaCysTyrAspIleThrIleAsp 190  
Qy 547 GGAGTATCTTTTCATCATGATGTCAGGATATAGGACAGATCTTATCAGTATTTT 606  
Db 191 SerMetProPheSerProTyrIleCysAlaGlyIleGlySerAspLeuValSerMetPhe 210  
Qy 607 AAAGACCTCAATCTAAAATTTGCTTACCAAGGAAAATAGGTATTAGTTACCTTATCACA 666  
Db 211 GluThrAsnProLysIleSerTyrGlnGlyLysLeuGlyValSerTyrSerIleSer 230  
Qy 667 CAGAGATCTCTGCATTTATTTGTTGATGATGATGATGATGATGATGATGATGATGATG 726  
Db 231 ProGluAlaSerValPheValGlyGlyHisPheHisArgValIleGlyAsnGluPheLys 250  
Qy 727 AAGATACCTGATTAATCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 786  
Db 251 AspIleProAlaIleThrProAlaGlyAlaThrGluIleLysGlyThrGlnPheThrThr 270  
Qy 787 GTAACTCTTGACGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840  
Db 271 ValThrLeuAsnIleCysHisPheGlyLeuGluLeuGlyGlyArgPheThrPhe 288

## RESULT 23

US-10-059-964-32

; Sequence 32, Application US/10059964					
; Patent No. US20020120115A1					
; GENERAL INFORMATION:					
; APPLICANT: Rikihisa, Yasuko					
; APPLICANT: Ohasi, No. US20020120115A1io					
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia					
; FILE REFERENCE: Chaffeensis					
; CURRENT APPLICATION NUMBER: US/10/059,964					
; CURRENT FILING DATE: 2002-01-28					
; EARLIER APPLICATION NUMBER: 09/314,701					
; EARLIER FILING DATE: 1999-05-19					
; NUMBER OF SEQ ID NOS: 66					
; SOFTWARE: PatentIn Ver. 2.0					
; SEQ ID NO 32					
; LENGTH: 288					
; TYPE: PRT					
; ORGANISM: Ehrlichia canis					
US-10-059-964-32					
Alignment Scores:					
Pred. No.:	1,23e-58	Length:	288		
Score:	621.00	Matches:	135		
Percent Similarity:	62.08%	Conservative:	50		
Best Local Similarity:	45.30%	Mismatches:	85		
Query Match:	41.93%	Indels:	28		
DB:	12	Gaps:	7		
US-10-062-624-41 (1-840) x US-10-059-964-32 (1-288)					
Qy	1	ATGAATTATAAGAAAATTTCTAGTAAGAAGCGGTTAAATCTCAATTAATGTCTCAATCTTACCA	60		
Db	1	MetAsnCysAlaYstYrMetProSerAlaSerAlaLeuLeuSerLeuMetSerPheLeuPro	20		
Qy	61	TATCAGTCCTTTCCAGATCCTGTAGTTCAGAACCTAATGATACAAGAAGGC---TTC	117		
Db	21	SerValSerPheSerGluSerIle-----HisGluAspAsnIleasnGlyAsnPhe	37		
Qy	118	TACATTAGTGCAAGTACCAATCCAAGTATATCACACTTTTAGAAAAATTTCTCTCTCAAGAA	177		
Db	38	TyrIleSerAlaYstYrMetProSerAlaSerHisPheGlyValPheSerVallysGlu	57		
Qy	178	ACTCCTATTAAATGGACAAATTTCTCTCAATAAAAGTTTTCCGACTAAAGAAGATGGT	237		
Db	58	-----GluYlsAsnThrThrGlyValPheGlyLeuYlsGlnAspTrp	72		
Qy	238	GAT---ATACAAAAAAGACGAT	261		
Db	73	ASPGlyAlaThrIleYlsAspAlaSerSerHisThrIleAspProSerThrIlePhe	92		
Qy	262	ACAAGAGTAGTCCAGGCATTTGATTTTCAAATAACTTAATATCATCAGGATTTTCAGGAAGT	321		
Db	93	SerIleSerAsnTySerPheYlStYrGluAsnAsnProPheLeuGlyPheAlaGlyAla	112		
Qy	322	ATTGGTTACTTATGACGCCAAGAAATAGACTTGAAGCTGCATATCAACAATTAAT	381		
Db	113	IleGlyTYrSerMetGlyGlyProArgValGluPheGluValSerTYrGluIlePheAsp	132		
Qy	382	CCAAAAACCACCGATACATGATCTAGTAATATGGTGAATATATATAACATTTTCGATTA	441		
Db	133	ValYlsAsnGlnGlyAsnSerTYrYlsAsnAspAla-----HisLYSTYrCysAlaLeu	150		
Qy	442	TCTCGT-----AAAGATGCATGGGAAGATCAGCAATATAGTACTTATAA	486		
Db	151	SerArgHisThrGlyGlyMetProGlnAlaGlyHisGlnAsnLysPheValPheLeuYls	170		
Qy	487	AATGACGGCATACTTTTATGTCATTGATGGTTAAATCTTGTATGACATTCACGCTGAA	546		
Db	171	AsnGlyGlyLeuLeuAspIleSerLeuMelleAsnAlaCystYrAspIleThrIleAsp	190		
Qy	547	GGAGTATCTTTTCGATACATATGATGTCAGGTPATAGGACGATCTTATCACTATTTT	606		
Db	191	SerMetProPheSerProTYrIleCysAlaGlyIleGlySerAspLeuValSerMetPhe	210		





```
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US/09/660,587
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,359
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 12
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1D
US-10-062-920-12

Alignment Scores:
Pred. No.: 1.57e-58 Length: 286
Score: 620.00 Matches: 139
Percent Similarity: 61.56% Conservative: 42
Best Local Similarity: 47.28% Mismatches: 89
Query Match: 41.86% Indels: 24
DB: 9 Gaps: 7

US-10-062-624-41 (1-840) x US-10-062-920-12 (1-286)
QY 1 ATGAATTATAAGAAATTCCTAGTAGAAGCGGTTAATCTCATTAATGCTCAATCTTACCA 60
DB 1 MetAsnCysGluLysPhePheIleThrAlaLeuThrLeuLeuMetSerPheLeuPro 20
QY 61 TATCAGTCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGATAACAAAGAGGC---TTC 117
DB 21 GlyIleSerLeuSerAspProVal-----GlnAspAspAsnIleSerGlyAsnPhe 37
QY 118 TACATTAGTCAAGTACAATCCAAAGTATATACACATTTAGAAAATTTCTGCTGAAGAA 177
DB 38 TyrIleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu 57
QY 178 ACTCCTATTAAATGGAACAAATTTCTCTCACTAAAAGTTTTCGGACTAAAGAAAGATGGT 237
DB 58 -----GluArgAsnThrThrValGlyValPheGlyIleGluGlnAspTrp 72
QY 238 GAT-----ATAACAAAAAGACGATTTTACAGAGTAGTCCAGGC 279
DB 73 AspArgCysValIleSerArgThrThrLeuSerAspIlePheThrValProAsnTyrSer 92
QY 280 ATTGATTTTCAAAATACTTAATATCAGGATTTTTCAGAGATATTTGTTACTCTATGGAC 339
DB 93 PheLysTyrGluAsnAsnLeuPheSerGlyPheAlaGlyAlaIleGlyTyrSerMetAsp 112
QY 340 GGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGATAAC 399
DB 113 GlyProArgIleGluLeuGluValSerTyrGluAlaPheAspValLysAsnGlnGlyAsn 132
QY 400 AATGATACTGATAATGTTGTAATCTATATAAATTTTGCATTATCTCGTAAAGATGCAATG 459
DB 460 GAAGATCAG-----CAATATGATAGTACTTAAATAATGACGCGATAACT 501
DB 151 GluThrGlnIleAspGlyAlaGlySerAlaSerValPheLeuIleAsnGluGlyLeu 170
QY 502 TTTATGTCATTGATGTTAACTTGTATGACATTACAGCTGAAGGAGTAGTCTTTCGTA 561
DB 171 AspLysSerPheMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSer 190
QY 562 CCATATGATGTCAGGATAGGACAGATCTTATCACTATTTTAAAGACCTCAATCTA 621
DB 191 ProTyrIleCysAlaGlyIleGlyIleAspLeuValSerMetPheGluAlaIleAsnPro 210
QY 622 AAATTTGCTTACCAGGAAAAATAGGTATTAGTACCCTATCACACCGAAGTCTCTGCA 681
DB 211 LysIleSerTyrGlnGlyLysLeuGlyLeuSerTyrProIleSerProGluAlaSerVal 230
QY 682 TTTATTTGGTGACTACTACCATGGCGGTTATTGGTAATAAATTTGAGAAGATACCTGTAATA 741
```

```
Db 231 PheIleGlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrMet 250
QY 742 ACTCTCTGTAGTATTAAATGATGCTCTCAA---ACCACATCTGCTTCAAGTAACCTTGAC 798
DB 251 IleProSerGluSerAlaLeuAlaGlyLysGlyAsnTyrProAlaIleValThrLeuAsp 270
QY 799 GTTGATACCTTTGCGGAGAAATTTGGAATGAGGTTCACCTTC 840
DB 271 ValPheTyrPheGlyIleGluLeuGlyGlyArgPheAsnPhe 284

RESULT 27
US-10-314-639-8
; Sequence 8, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20030103991A1
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/314,701
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-10-314-639-8

Alignment Scores:
Pred. No.: 1.57e-58 Length: 286
Score: 620.00 Matches: 139
Percent Similarity: 61.56% Conservative: 42
Best Local Similarity: 47.28% Mismatches: 89
Query Match: 41.86% Indels: 24
DB: 9 Gaps: 7

US-10-062-624-41 (1-840) x US-10-314-639-8 (1-286)
QY 1 ATGAATTATAAGAAATTCCTAGTAGAAGCGGTTAATCTCATTAATGCTCAATCTTACCA 60
DB 1 MetAsnCysGluLysPhePheIleThrAlaLeuThrLeuLeuMetSerPheLeuPro 20
QY 61 TATCAGTCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGATAACAAAGAGGC---TTC 117
DB 21 GlyIleSerLeuSerAspProVal-----GlnAspAspAsnIleSerGlyAsnPhe 37
QY 118 TACATTAGTCAAGTACAATCCAAAGTATATACACATTTAGAAAATTTCTGCTGAAGAA 177
DB 38 TyrIleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu 57
QY 178 ACTCCTATTAAATGGAACAAATTTCTCTCACTAAAAGTTTTCGGACTAAAGAAAGATGGT 237
DB 58 -----GluArgAsnThrThrValGlyValPheGlyIleGluGlnAspTrp 72
QY 238 GAT-----ATAACAAAAAGACGATTTTACAGAGTAGTCCAGGC 279
DB 73 AspArgCysValIleSerArgThrThrLeuSerAspIlePheThrValProAsnTyrSer 92
QY 280 ATTGATTTTCAAAATACTTAATATCAGGATTTTTCAGAGATATTTGTTACTCTATGGAC 339
DB 93 PheLysTyrGluAsnAsnLeuPheSerGlyPheAlaGlyAlaIleGlyTyrSerMetAsp 112
QY 340 GGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGATAAC 399
DB 113 GlyProArgIleGluLeuGluValSerTyrGluAlaPheAspValLysAsnGlnGlyAsn 132
QY 400 AATGATACTGATAATGTTGTAATCTATATAAATTTTGCATTATCTCGTAAAGATGCAATG 459
DB 460 GAAGATCAG-----CAATATGATAGTACTTAAATAATGACGCGATAACT 501
DB 151 GluThrGlnIleAspGlyAlaGlySerAlaSerValPheLeuIleAsnGluGlyLeu 170
QY 502 TTTATGTCATTGATGTTAACTTGTATGACATTACAGCTGAAGGAGTAGTCTTTCGTA 561
DB 171 AspLysSerPheMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSer 190
QY 562 CCATATGATGTCAGGATAGGACAGATCTTATCACTATTTTAAAGACCTCAATCTA 621
DB 191 ProTyrIleCysAlaGlyIleGlyIleAspLeuValSerMetPheGluAlaIleAsnPro 210
QY 622 AAATTTGCTTACCAGGAAAAATAGGTATTAGTACCCTATCACACCGAAGTCTCTGCA 681
DB 211 LysIleSerTyrGlnGlyLysLeuGlyLeuSerTyrProIleSerProGluAlaSerVal 230
QY 682 TTTATTTGGTGACTACTACCATGGCGGTTATTGGTAATAAATTTGAGAAGATACCTGTAATA 741
```

Db 133 Asn-----TyrLysAsnGluAlaHisArgTyrTyrAlaLeuSerHisLeuLeuGlyThr 150  
Qy 460 GAAGATCAG-----CAATATGTAGTACTTAAATAATGACGGCATAACT 501  
Db 151 GluThrGlnIleAspGlyAlaGlySerAlaSerValPheLeuIleAsnGluLeu 170  
Qy 502 TTTATGTCATTGATGGTTAACTTCTGATACATTACAGTGAAGGATGATCTTTCGTA 561  
Db 171 AspLysSerPheMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSer 190  
Qy 562 CCATATGATGTGAGGATAGGAGCAGATCTTACTTATTTTAAAGACCTCAATCTA 621  
Db 191 ProTyrIleCysAlaGlyIleGlyLeuAspLeuValSerMetPheGluAlaIleAsnPro 210  
Qy 622 AAATTTGCTTACCAAGGAAATAGTATTAGTTACCCCTATCACACCAAGAGTCTCTGCA 681  
Db 211 LysIleSerTyrGlnGlyLysLeuGlyLeuSerTyrProIleSerProGluAlaSerVal 230  
Qy 682 TTTATTGGTGATACTACCATGGGCTTATTCGTTAATAATTTGAGAGATACCTGTAATA 741  
Db 231 PheIleGlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrMet 250  
Qy 742 ACTCTGTAGTATTAAATGATGCTCTCAAA--ACCACATCTGCTTCAGTAACTCTTGAC 798  
Db 251 IleProSerGluSerAlaLeuAlaGlyLysGlyAsnTyrProAlaIleValThrLeuAsp 270  
Qy 799 GTTGGATCTTTGGCGGAGAAATTTGGAATGAGTTCACCTTC 840  
Db 271 ValPheTyrPheGlyIleGluLeuGlyGlyArgPheAsnPhe 284

## RESULT 28

US-09-846-808-16  
; Sequence 16, Application US/09846808  
; Patent No. US20020064531A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: Yu, Xu-Jie  
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane  
; FILE REFERENCE: Protein Multigene Family  
; CURRENT APPLICATION NUMBER: US/09/846,808  
; CURRENT FILING DATE: 2001-05-01  
; PRIOR APPLICATION NUMBER: 60/201,035  
; PRIOR FILING DATE: 2000-05-01  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 16  
; LENGTH: 286  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: P28-16 Outer Membrane Protein of  
; OTHER INFORMATION: Ehrlichia chaffeensis

US-09-846-808-16

Alignment Scores:  
Pred. No.: 1.57e-58 Length: 286  
Score: 620.00 Matches: 139  
Percent Similarity: 61.56% Conservative: 42  
Best Local Similarity: 47.28% Mismatches: 89  
Query Match: 41.86% Indels: 24  
DB: 10 Gaps: 7

US-10-062-624-41 (1-840) x US-09-846-808-16 (1-286)

Qy 1 ATGAATTATAGAAATTTAGTAGAAGCGCGTTAATCTCATTAAATGTCATCTTACCA 60  
Db 1 MetAsnCysGluLysPhePheIleThrThrAlaLeuThrLeuLeuMetSerPheLeuPro 20  
Qy 61 TATCAGTCTTTTTCAGATCTCTGTAGTTCAGAACTAATGATCAACAAAGAGGCG--TTC 117  
Db 21 GlyIleSerLeuSerAspProVal-----GlnAspAspAsnIleSerGlyAsnPhe 37  
Qy 118 TACATTAGTCAAAAGTACAATCCAAAGTATATACACTTTAGAAAAATTTCTCTGCTGAAGAA 177

Db 38 TyrIleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu 57  
Qy 178 ACTCTATTATGAACCAAAATTTCTCTCACTAAAAAGTTTTTCGACTATAAGAAAGATGGT 237  
Db 58 -----GluArgAsnThrThrValGlyValPheGlyIleGluGlnAspTrp 72  
Qy 238 GAT-----ATAACAAAAAGACAGATTTTACAAGAGTAGCTCCAGGC 279  
Db 73 AspArgCysValIleSerArgThrThrLeuSerAspIlePheThrValProAsnTyrSer 92  
Qy 280 ATTGATTTTCAAAATACTTAAATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGAC 339  
Db 93 PheLysTyrGluAsnAsnLeuPheSerGlyPheAlaGlyAlaIleGlyTyrSerMetAsp 112  
Qy 340 GGACCAAGATAGAACCTTGAAGCTGCATATCAACAATTTTAAATCCAAAAACACCGATAAC 399  
Db 113 GlyProArgIleGluLeuGluValSerTyrGluAlaPheAspValLysAsnGlnGlyAsn 132  
Qy 400 AATGATACTGATTAATGCTGAATATCAATATTTGTCATTATCTCGTAAAGATGCAATG 459  
Db 133 Asn-----TyrLysAsnGluAlaHisArgTyrTyrAlaLeuSerHisLeuLeuGlyThr 150  
Qy 460 GAAGATCAG-----CAATATGCTAGTACTTAAATAATGACGGCATAACT 501  
Db 151 GluThrGlnIleAspGlyAlaGlySerAlaSerValPheLeuIleAsnGluGlyLeu 170  
Qy 502 TTTATGTCATTGATGGTTAATCTGCTATGACATTACAGCTGAAGGAGTATCTTTCGTA 561  
Db 171 AspLysSerPheMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSer 190  
Qy 562 CCATATGCTGTCAGGTATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTA 621  
Db 191 ProTyrIleCysAlaGlyIleGlyIleAspLeuValSerMetPheGluAlaIleAsnPro 210  
Qy 622 AAATTTGCTTACCAAGGAAATAGTATTAGTTACCTTATCACACCAAGAGTCTCTGCA 681  
Db 211 LysIleSerTyrGlnGlyLysLeuGlyLeuSerTyrProIleSerProGluAlaSerVal 230  
Qy 682 TTTATTGGTGATACTACCATGGCTTATTTGTTAATAATTTGAGAGATACCTGTAATA 741  
Db 231 PheIleGlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrMet 250  
Qy 742 ACTCTGTAGTATTAAATGATGCTCTCAAA--ACCACATCTGCTTCAGTAACTCTTGAC 798  
Db 251 IleProSerGluSerAlaLeuAlaGlyLysGlyAsnTyrProAlaIleValThrLeuAsp 270  
Qy 799 GTTGGATCTTTGGCGGAGAAATTTGGAATGAGTTCACCTTC 840  
Db 271 ValPheTyrPheGlyIleGluLeuGlyGlyArgPheAsnPhe 284

## RESULT 29

US-10-062-994-15  
; Sequence 15, Application US/10062994  
; Patent No. US20010001661A1  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Bowie, Michael V.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Mahan, Suman M.  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Moreland, Annie L.  
; APPLICANT: Simbi, Bigboy H.  
; APPLICANT: Whitmore, William M.  
; APPLICANT: Allenman, Arthur R.  
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and  
; FILE REFERENCE: Methods of use  
; FILE REFERENCE: UF-167XC3  
; CURRENT APPLICATION NUMBER: US/10/062,994  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: 09/533,662  
; PRIOR FILING DATE: 2000-04-21

```
; PRIOR APPLICATION NUMBER: 09/337,827
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 08/953,326
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 08/733,230
; PRIOR FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patent in ver. 2.0
; SEQ ID NO 15
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-10-062-994-15

Alignment Scores:
Pred. No.: 1.57e-58 Length: 286
Score: 620.00 Matches: 139
Percent Similarity: 61.56% Conservative: 42
Best Local Similarity: 47.28% Mismatches: 89
Query Match: 41.86% Indels: 24
DB: 12 Gaps: 7

US-10-062-624-41 (1-840) x US-10-062-994-15 (1-286)
Qy 1 ATGAATTATAGAAATTCCTAGTAAGAGCGGTTAAATCTCATTAATCTCAATCTTACCA 60
Db 1 MetAsnCysGluLysPhePheIleThrAlaLeuThrLeuLeuMetSerPheLeuPro 20
Qy 61 TATCAGTCTTTGTCAGATCCTGTAGGTTCAAGAACTAATGATAACAAAGAGGC---TTC 117
Db 21 GlyLeuSerLeuSerAspProVal-----GlnAspAspAsnIleSerGlyAsnPhe 37
Qy 118 TACATTAGTCAAAATCAATCCAAAGTATATACACTTTAGAAAATTCCTGCTGAAGAA 177
Db 38 TyrIleSerGlyLysTyMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu 57
Qy 178 ACTCCTATTATGGAACAATTCCTCACTAAAAAGTTTCGGACTAAAGAAAGATGGT 237
Db 58 -----GluArgAsnThrValGlyValPheGlyLeuGlnAsnTrp 72
Qy 238 GAT-----ATAACAAAAAGACGATTTTACAGAGTAGTAGTCCAGGC 279
Db 73 AspArgCysValIleSerArgThrThrLeuSerAspIlePheThrValProAsnTyrSer 92
Qy 280 ATTGATTTTCAAAATCAATTAATATCAGAGATTTTTCAGGAAGTATTTGGTTACTTATGAC 339
Db 93 PheLysTyrGluAsnLeuPheSerGlyPheAlaGlyAlaIleGlyTyrSerMetAsp 112
Qy 340 GGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGATAC 399
Db 113 GlyProArgIleGluLeuGluValSerTyrGluAlaPheAspValLysAsnGlnGlyAsn 132
Qy 400 AATGATCTGATAATGGTGAATCTTCAATAACATTTTGCATTTCTCGTAAAGATCAATG 459
Db 133 Asn-----TyrLysAsnGluAlaHisArgTyrTyrAlaLeuSerHisLeuLeuGlyThr 150
Qy 460 GAAGATCAG-----CAATATGATAGTACTTAAATAATGACGGCATAACT 501
Db 151 GluThrGlnIleAspGlyAlaGlySerAlaSerValPheLeuIleAsnGlnGlyLeu 170
Qy 502 TTTATGTCATGATGTTAATCTTGTATGATTCATACAGTGAAGAGTAGTCTTTTCGTA 561
Db 171 AspLysSerPheMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSer 190
Qy 562 CCATATGATGTCAGGATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTA 621
Db 191 ProTyrIleCysAlaGlyIleGlyLeuAspLeuValSerMetPheGluAlaIleAsnPro 210
Qy 622 AATTTGCTTACCAAGGAAAAATAGGTATTAGTTACCTTATCACCACACAGAGATCTCTGCA 681
Db 211 LysIleSerTyrGlnGlyLysLeuGlyLeuSerTyrProIleSerProGluAlaSerVal 230
Qy 682 TTTATTGGTGGATACCTACCATCGCGTTATTGGTAATAATTTGAGAAGATACCTGTAATA 741
```

```
Db 231 PheIleGlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrMet 250
Qy 742 ACTCCTCTAGTATTAAATGATGCTCTCAAA---ACCACATCTGCTTCACTAATCTTGGAC 798
Db 251 IleProSerGluSerAlaLeuAlaGlyLysGlyAsnTyrProAlaIleValThrLeuAsp 270
Qy 799 GTTGATACACTTTGGCGAGAAATTTGGAATGAGGTTCACCTTC 840
Db 271 ValPheTyrPheGlyIleGluLeuGlyGlyArgPheAsnPhe 284

RESULT 30
US-10-062-624-12
; Sequence 12, Application US/10062624
; Patent No. US20020115840A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-Kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2/D1
; CURRENT APPLICATION NUMBER: US/10/062,624
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 09/660,587
; PRIOR FILING DATE: 2000-09-12
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 12
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1D
US-10-062-624-12

Alignment Scores:
Pred. No.: 1.57e-58 Length: 286
Score: 620.00 Matches: 139
Percent Similarity: 61.56% Conservative: 42
Best Local Similarity: 47.28% Mismatches: 89
Query Match: 41.86% Indels: 24
DB: 12 Gaps: 7

US-10-062-624-41 (1-840) x US-10-062-624-12 (1-286)
Qy 1 ATGAATTATAGAAATTCCTAGTAAGAGCGGTTAAATCTCATTAATCTCAATCTTACCA 60
Db 1 MetAsnCysGluLysPhePheIleThrAlaLeuThrLeuLeuMetSerPheLeuPro 20
Qy 61 TATCAGTCTTTGTCAGATCCTGTAGGTTCAAGAACTAATGATAACAAAGAGGC---TTC 117
Db 21 GlyLeuSerLeuSerAspProVal-----GlnAspAspAsnIleSerGlyAsnPhe 37
Qy 118 TACATTAGTCAAAATCAATCCAAAGTATATACACTTTAGAAAATTCCTGCTGAAGAA 177
Db 38 TyrIleSerGlyLysTyMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu 57
Qy 178 ACTCCTATTATGGAACAATTCCTCACTAAAAAGTTTCGGACTAAAGAAAGATGGT 237
Db 58 -----GluArgAsnThrThrValGlyValPheGlyLeuGlnAsnTrp 72
Qy 238 GAT-----ATAACAAAAAGACGATTTTACAGAGTAGTAGTCCAGGC 279
Db 73 AspArgCysValIleSerArgThrThrLeuSerAspIlePheThrValProAsnTyrSer 92
Qy 280 ATTGATTTTCAAAATCAATTAATATCAGAGATTTTTCAGGAAGTATTTGGTTACTTATGAC 339
Db 93 PheLysTyrGluAsnLeuPheSerGlyPheAlaGlyAlaIleGlyTyrSerMetAsp 112
Qy 340 GGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGATAC 399
Db 113 GlyProArgIleGluLeuGluValSerTyrGluAlaPheAspValLysAsnGlnGlyAsn 132
Qy 400 AATGATCTGATAATGGTGAATCTTCAATAACATTTTGCATTTCTCGTAAAGATCAATG 459
Db 133 Asn-----TyrLysAsnGluAlaHisArgTyrTyrAlaLeuSerHisLeuLeuGlyThr 150
Qy 460 GAAGATCAG-----CAATATGATAGTACTTAAATAATGACGGCATAACT 501
Db 151 GluThrGlnIleAspGlyAlaGlySerAlaSerValPheLeuIleAsnGlnGlyLeu 170
Qy 502 TTTATGTCATGATGTTAATCTTGTATGATTCATACAGTGAAGAGTAGTCTTTTCGTA 561
Db 171 AspLysSerPheMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSer 190
Qy 562 CCATATGATGTCAGGATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTA 621
Db 191 ProTyrIleCysAlaGlyIleGlyLeuAspLeuValSerMetPheGluAlaIleAsnPro 210
Qy 622 AATTTGCTTACCAAGGAAAAATAGGTATTAGTTACCTTATCACCACACAGAGATCTCTGCA 681
Db 211 LysIleSerTyrGlnGlyLysLeuGlyLeuSerTyrProIleSerProGluAlaSerVal 230
Qy 682 TTTATTGGTGGATACCTACCATCGCGTTATTGGTAATAATTTGAGAAGATACCTGTAATA 741
```

QY		400	AATGATCTAGTAAATGGTGGAATACTATAAACATTMTTCGCATTATCTCGTTAAGAAGTCAATG	459
			: : : : :	
DB		133	Asn-----Ty rLysAsnGluAlaHisArgTyrTy rAlaLeuSerHisLeuLeuGlyThr	150
			: : : : :	
QY		460	GAAGATCAG-----CAATATGTACTACTTAAAATACGCCGCAATACT	501
			: : : : :	
DB		151	GluThrGlnIleAspGlyAlaGlySerAlaSerValPheLeuleahenGluGlyLeuLeu	170
			: : : : :	
QY		502	TTTATGTCAATCATGGTTTAATCTTGTCTATCACATTCAGAGGAGTATCTTCGTA	561
			: : : : :	
DB		171	AspLysSerPheMetLeuAsnAlaCysTy rAspValIleSerGluGlyIleProPheSer	190
			: : : : :	
QY		562	CCATATGCATGTGCAGGTATAGGACAGATCTTATCACTATTTTTAAAGACCTCAATCTA	621
			: : : : :	
DB		191	ProTyrIleCysAlaGlyIleGlyIleAspleuValSerMetPheGluAlaIleasnPro	210
			: : : : :	
QY		622	AAATTTGCTTACCAGGAAAAATAGGTATTAGTTACCTATCACACGAGAAGTCTCGCA	681
			: : : : :	
DB		211	LysIleSerTy rGinGlyLysLeuGlyLeuSerTy rProIleSerProGluAlaSerVal	230
			: : : : :	
QY		682	TTTATTGTGTGATACTACCATGGCGTTATTGGTAAATAATTTGAGAAGATACCTGTATA	741
			: : : : :	
DB		231	PheIleGlyGlyHisPheHisLysValIleGlyAsnGluPheargAspileproThrMet	250
			: : : : :	
QY		742	ACTCCCTGTAGTATTAAATGATGCTTCCTCAA---ACCACATCTCCTTCAGTAACTCTTGAC	798
			: : : : :	
DB		251	IleProSerGluSerAlaLeuAlaGlyIleGlyAsnTy rProAlaIleValThrIleuasP	270
			: : : : :	
QY		799	GTTGGATCTATTGGCGGAGAAATTGGAATCAGGTTCCACCTTC	840
			: : : : :	
DB		271	ValPheTy rPheGlyIleGluLeuGlyGlyVarGAPheAsnPhe	284
			: : : : :	

RESULT 31

```

US-10-059-964-8
; Sequence 8, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20020120115A110
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; CURRENT FILING DATE: 2002-01-28
; EARLIER APPLICATION NUMBER: 09/314,701
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 8
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-10-059-964-8

```

Alignment Scores:	
Pred. NO.:	1.57e-58
Score:	620.00
Length:	286
Matches:	139
Percent Similarity:	61.56%
Conservative:	42
Best Local Similarity:	47.28%
Mismatches:	89
Query Match:	41.86%
Indels:	24
Gaps:	7
DB:	12

US-10-062-624-41 (1-840) x US-10-059-964-8 (1-286)

Qy	1	ATGATTATAAGAAAATTC	TAGTAGAAGCGGTTAAT	TCTCAATTAATGCAATCTTACCA	60
			:::		
Db	1	MetAenCysGluLysPhe	PheIleThrAlaLeuThrLeu	LeuMetSerPheLeuPro	20
		:::	:::		
Qy	61	TATCAGCTCTTTGAGATCT	TGTAGGTTCAAGAACTAAT	GATACAAAGAACG---	TTC 117
			:::		
Db	21	GlylleSerLeuSerAppro	Val-----GlnAspAsn	IleSerGlyAsnPhe	37
		:::	:::		
Qy	118	TACATTAGTGCAAAGTACA	TTCACCAAGTATATCACT	TTTAGAAAATTCCTGCTGAAGA	177

38	Tyr	Ile	Ser	Gly	Ile	Val	Met	Pro	Ser	Ala	Ser	His	Phe	Gly	Val	Phe	Ser	Ala	Ile	Ser	Glu	57
178	ACT	CCT	TAT	TAA	TGCA	CAAA	AAT	TCT	CTC	ACT	AAAA	AGT	TTT	CGG	ACT	AAAG	AAAC	GAT	GGT	237		
58	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	72	
238	GAT	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	279	
73	Asp	Arg	Cys	Val	Ile	Ser	Arg	Thr	Thr	Leu	Ser	Asp	Ile	Phe	Thr	Val	Pro	Asn	Tyr	Ser	92	
280	ATT	GAT	TTT	CAA	AAT	AACT	TAA	TAT	CAG	GAT	TTT	CAG	GAAG	TAT	TGG	TACT	CTC	TAT	GGAC	339		
93	Phe	Ile	Tyr	Glu	Asn	Asn	Leu	Phe	Ser	Gly	Phe	Ala	Gly	Ala	Ile	Gly	Tyr	Ser	Met	Asp	112	
340	GG	ACC	AGA	GAT	TAG	AACT	TGG	AGCT	GCA	TAT	CA	CAAT	TTT	AA	TCC	AAAA	AAAC	ACC	GAT	TAAC	399	
113	Gly	Pro	Arg	Ile	Glu	Leu	Glu	Val	Ser	Tyr	Glu	Ala	Phe	Asp	Val	Ile	Asn	Gln	Gly	Asn	132	
400	AAT	GAT	ACT	GAT	AAT	TGG	TGA	TACT	ATA	AAAC	ATTT	TGC	AAT	TAT	CTC	GTA	AAAG	ATC	CAAT	G	459	
133	Asn	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	150	
460	GA	GAT	CAG	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	501	
151	Glu	Thr	Gln	Ile	Asp	Gly	Ile	Ala	Gly	Ser	Ala	Ser	Val	Phe	Ile	Asn	Glu	Gly	Leu	Leu	170	
502	TTT	ATG	CAT	TGAT	TGG	TTA	TACT	TGCT	ATG	CATT	TAC	AGCT	GAAG	GAT	TAT	CTT	CTG	TA	561			
171	Asp	Ile	Ser	Ser	Phe	Met	Leu	Asn	Ala	Cys	Tyr	Asp	Val	Ile	Ser	Glu	Gly	Ile	Pro	Phe	190	
562	CCAT	ATG	CGAT	TGCG	AGT	TAT	GAG	CAG	AGAT	CTT	TAT	CAT	TAT	TTT	TTA	AGAC	CTC	TAAT	CTA	621		
191	Pro	Tyr	Ile	Cys	Ala	Gly	Ile	Gly	Ile	Asp	Leu	Val	Ser	Met	Phe	Glu	Ala	Ile	Asn	Pro	210	
622	AAAT	TTCT	CTT	ACCA	GGAAAA	TAG	TAT	TAG	TAT	TAG	TAC	CTT	ATC	CTT	CAT	CAC	ACC	AGAG	CTCT	CTCA	681	
211	Lys	Ile	Ser	Tyr	Gln	Gly	Ile	Ser	Gly	Leu	Ser	Tyr	Pro	Ile	Ser	Pro	Glu	Ala	Ser	Val	230	
682	TTT	ATT	GGT	GGAT	CTAC	CACT	ACCA	TGCG	TTT	TGG	TAA	TAA	TTT	GAGA	AGAT	CACT	CTG	TAATA	741			
231	Phe	Ile	Gly	Gly	His	Phe	His	Ile	Val	Ile	Gly	Asn	Glu	Phe	Arg	Asp	Ile	Pro	Thr	Met	250	
742	ACT	CCT	G	TAG	TAT	TAA	ATG	ATG	CTC	CTCA	---	---	---	---	---	---	---	---	---	---	798	
251	Ile	Pro	Ser	Glu	Ser	Ala	Leu	Ala	Gly	Ile	Ser	Gly	Asn	Tyr	Pro	Ala	Ile	Val	Thr	Leu	Asp	270
799	GTT	GGAT	CTAT	TTT	GGCG	GAG	AAAT	TGG	AAAT	TGG	AAAT	TGG	AAAT	TGG	AAAT	TGG	AAAT	TGG	AAAT	TGG	840	
271	Val	Phe	Tyr	Phe	Gly	Ile	Glu	Leu	Gly	Ile	Val	Arg	Phe	Asn	Phe	284						

RESULT 32

```

US-062-994-15
; Sequence 15, Application US/10062994
; Patent No. US20020132789A1
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Bowie, Michael V.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Mahan, Suman M.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Kurangirwa, Fred R.
; APPLICANT: Moreland, Annie L.
; APPLICANT: Simbi, Bigboy H.
; APPLICANT: Whitmire, William M.
; APPLICANT: Alteman, Arthur R.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and
; TITLE OF INVENTION: Methods of use
; FILE REFERENCE: UF-167XC3
; CURRENT APPLICATION NUMBER: US/10/062,994
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 09/533,662
; PRIOR FILING DATE: 2000-04-21

```

```
; PRIOR APPLICATION NUMBER: 09/337,827
; PRIOR FILING DATE: 1999-08-22
; PRIOR APPLICATION NUMBER: 08/953,326
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 08/733,230
; PRIOR FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 15
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-10-062-994-15

Alignment Scores:
Pred. No.: 1,576-58 Length: 286
Score: 620.00 Matches: 139
Percent Similarity: 61.56% Conservative: 42
Best Local Similarity: 47.28% Mismatches: 89
Query Match: 41.86% Indels: 24
DB: 12 Gaps: 7

US-10-062-624-41 (1-840) x US-10-062-994-15 (1-286)

QY 1 ATGAATTATAGAAAATTCAGTAAAGACGGGTTAATCTCATTAATGCTCAATCTTACCA 60
DB 1 MetAsnCysGluLeuPhePheIleThrAlaLeuThrLeuLeuMetSerPheLeuPro 20
QY 61 TATCAGTCTTTTGCAGATCCTCTAGGTTCAAGAACTAATGATAACAAAGAGGC---TTC 117
DB 21 GlyIleSerLeuSerAspVal-----GlnAspAsnIleSerGlyAsnPhe 37
QY 118 TACATTAGTGCAGAACTACCAATCAAGTATATCACACTTTAGAAAATCTCTGCTGAAGAA 177
DB 38 TyrIleSerGlyLeuTyrosMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu 57
QY 178 ACTCCTATTATGGAACAATCTCTCACTAAAAAGTTTCGGACTAAAGAAAGATGGT 237
DB 58 -----GluArgAsnThrThrValGlyValPheGlyIleGluGlnAspTrp 72
QY 238 GAT-----ATACAAAAAAGACGATTTTACAAAGAGTAGTCCAGGC 279
DB 73 AspArgCysValIleSerArgThrThrLeuSerAspIlePheThrValProAsnTyrSer 92
QY 280 ATTGATTTTCAAAATTAATATACAGATTTTCAAGAGATTGTTGTTACTCTATGAC 339
DB 93 PheIleTyrGluAsnLeuPheSerGlyPheAlaGlyAlaIleGlyTyrSerMetAsp 112
QY 340 CGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGATAC 399
DB 113 GlyProA-gilleGluLeuGluValSerTyrGluAlaPheAspValIleAsnGlnGlyAsn 132
QY 400 AATGATACTGATAATGGTGAATCTTCACTAAACATTTTGGCATTTCTCGTAAAGATCAATG 459
DB 133 Asn-----TyrLysAsnGluAlaHisArgTyrTyrAlaLeuSerHisLeuLeuGlyThr 150
QY 460 GAAGATCAG-----CAATATGTAGTACTTTAAATAATGACGGCATAACT 501
DB 151 GluThrGlnIleAspGlyAlaGlySerAlaSerValPheLeuIleAsnGluGlyLeu 170
QY 502 TTTATGTCATTGATGTTAATCTTGTATGATACATACAGCTGAAGGAGTATCTTCGTA 561
DB 171 AspLysSerPheMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSer 190
QY 562 CCATATGATGTCAGGTATAGGACAGATCTTATCACTATTATTTTAAAGACCTCAATCTA 621
DB 191 ProTyrIleCysAlaGlyIleGlyLeuAspLeuValSerMetPheGluAlaIleAsnPro 210
QY 622 AAATTTGCTTACCAAGGAAAAATAGTATTAGTTACCTTATCACACCGAAGTCTCTGCA 681
DB 211 LysIleSerTyrGlnGlyLysLeuGlyLeuSerTyrProIleSerProGluAlaSerVal 230
QY 682 TTTATTGGTGGTACTACCATCGCGTTATTGGTAAATAATTTGAGAGATACCTGTAATA 741

; PRIOR APPLICATION NUMBER: 09/337,827
; PRIOR FILING DATE: 1999-08-22
; PRIOR APPLICATION NUMBER: 08/953,326
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 08/733,230
; PRIOR FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 15
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-10-062-994-15

Alignment Scores:
Pred. No.: 1,576-58 Length: 286
Score: 620.00 Matches: 139
Percent Similarity: 61.56% Conservative: 42
Best Local Similarity: 47.28% Mismatches: 89
Query Match: 41.86% Indels: 24
DB: 12 Gaps: 7

US-10-062-624-41 (1-840) x US-10-062-994-15 (1-286)

QY 1 ATGAATTATAGAAAATTCAGTAAAGACGGGTTAATCTCATTAATGCTCAATCTTACCA 60
DB 1 MetAsnCysGluLeuPhePheIleThrAlaLeuThrLeuLeuMetSerPheLeuPro 20
QY 61 TATCAGTCTTTTGCAGATCCTCTAGGTTCAAGAACTAATGATAACAAAGAGGC---TTC 117
DB 21 GlyIleSerLeuSerAspVal-----GlnAspAsnIleSerGlyAsnPhe 37
QY 118 TACATTAGTGCAGAACTACCAATCAAGTATATCACACTTTAGAAAATCTCTGCTGAAGAA 177
DB 38 TyrIleSerGlyLeuTyrosMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu 57
QY 178 ACTCCTATTATGGAACAATCTCTCACTAAAAAGTTTCGGACTAAAGAAAGATGGT 237
DB 58 -----GluArgAsnThrThrValGlyValPheGlyIleGluGlnAspTrp 72
QY 238 GAT-----ATACAAAAAAGACGATTTTACAAAGAGTAGTCCAGGC 279
DB 73 AspArgCysValIleSerArgThrThrLeuSerAspIlePheThrValProAsnTyrSer 92
QY 280 ATTGATTTTCAAAATTAATATACAGATTTTCAAGAGATTGTTGTTACTCTATGAC 339
DB 93 PheIleTyrGluAsnLeuPheSerGlyPheAlaGlyAlaIleGlyTyrSerMetAsp 112
QY 340 CGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGATAC 399
DB 113 GlyProA-gilleGluLeuGluValSerTyrGluAlaPheAspValIleAsnGlnGlyAsn 132
QY 400 AATGATACTGATAATGGTGAATCTTCACTAAACATTTTGGCATTTCTCGTAAAGATCAATG 459
DB 133 Asn-----TyrLysAsnGluAlaHisArgTyrTyrAlaLeuSerHisLeuLeuGlyThr 150
QY 460 GAAGATCAG-----CAATATGTAGTACTTTAAATAATGACGGCATAACT 501
DB 151 GluThrGlnIleAspGlyAlaGlySerAlaSerValPheLeuIleAsnGluGlyLeu 170
QY 502 TTTATGTCATTGATGTTAATCTTGTATGATACATACAGCTGAAGGAGTATCTTCGTA 561
DB 171 AspLysSerPheMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSer 190
QY 562 CCATATGATGTCAGGTATAGGACAGATCTTATCACTATTATTTTAAAGACCTCAATCTA 621
DB 191 ProTyrIleCysAlaGlyIleGlyLeuAspLeuValSerMetPheGluAlaIleAsnPro 210
QY 622 AAATTTGCTTACCAAGGAAAAATAGTATTAGTTACCTTATCACACCGAAGTCTCTGCA 681
DB 211 LysIleSerTyrGlnGlyLysLeuGlyLeuSerTyrProIleSerProGluAlaSerVal 230
QY 682 TTTATTGGTGGTACTACCATCGCGTTATTGGTAAATAATTTGAGAGATACCTGTAATA 741

; Sequence 14, Application US/10062051
; Publication No. US20030073095A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/10/062,051
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US/09/660,587
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 14
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F
US-10-062-051-14

Alignment Scores:
Pred. No.: 6,66-57 Length: 280
Score: 605.00 Matches: 133
Percent Similarity: 63.39% Conservative: 54
Best Local Similarity: 45.08% Mismatches: 78
Query Match: 40.85% Indels: 30
DB: 9 Gaps: 10

US-10-062-624-41 (1-840) x US-10-062-051-14 (1-280)

QY 1 ATGAATTATAGAAAATTCAGTAAAGACGGGTTAATCTCATTAATGCTCAATCTTACCA 60
DB 1 MetAsnCysGlyLysPhePheIleThrThrLeuValSerLeuMetSerPheLeuPro 20
QY 61 TATCAGTCTTTTGCAGATCCTCTAGGTTCAAGAACTAATGATAACAAAGAGGC---TTC 117
DB 21 GlyIleSerPheSerAspAlaVal-----GlnAsnAspAsnValGlyGlyAsnPhe 37
QY 118 TACATTAGTGCAGAACTACCAATCAAGTATATCACACTTTAGAAAATCTCTGCTGAAGAA 177
DB 38 TyrIleSerGlyLysTyrValProSerValSerHisPheGlyValPheSerAlaLysGln 57
QY 178 ACTCCTATTATGGAACAATCTCTCACTAAAAAGTTTTCGGACTAAAGAAAGATGGT 237
DB 58 -----GluArgAsnThrThrThrGlyValPheGlyLeuLysGlnAspTrp 72
QY 238 GAT-----ATACAAAAAAGAGC---GATTTTACAGAGTAGTCCAGGCATTGAT 285
DB 73 AspGlySerThrIleSerLysAsnSerProGluAsnThrPheAsnValProAsnTyrSer 92
QY 286 TTT-----CAAAATAACTTAATATCAGGATTTTCAGGAAGTATTTGTTACTCTATGAC 339
DB 93 PheLysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaValGlyTyrLeuMetAsn 112
QY 340 GGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGATAC 399
```

113	GlyProArgIleGluLeuGluMetSerTyrGluThrPheAspValIysAsnGlnGlyAsn	132
400	AATGATACTGATAAATGGTGAATACTATAACATATTTGCAATTATCTCGT-----	447
133	AsnTyrLysAsnAspAla-----HisLysTyrTyrAlaLeuThrHisAsnSerGlyGly	150
448	AAAGATGCAATGGAAGATCAGCAATATGCTAGTACTTAAATAATGACGCATAACTTTATG	507
151	LysLeuSerAsnAlaGlyAspLysPheValPheLeuLysAsnGluGlyLeuLeuAspIle	170
508	TCATTTGATGGTTAAATCACTTGCATGACATATACAGCTGAAGAGAGTATCTTTTCGTAACCATAT	567
171	SerLeuMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSerProTyr	190
568	GCATGTGCAGGTATAGGACGAGACTTATCACTATTTTAAAGACCTCAATCTAAATTT	627
191	IleCysAlaGlyValGlyThrAspLeuIleSerMetPheGluAlaIleAsnProLysIle	210
628	GCTTACCAAGGAAAATAGGTATTAGTTACCTCTATCACACAGAAAGTCTCGCATTTATT	687
211	SerTyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSerValPheVal	230
688	GGTGGATACCATCGGTATTGGTAAATAAATTTGAGAAGATACCTGTGAATAACTCTCT	747
231	GlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProAlaMetIlePro	250
748	GTAGTATTAAATGATGCTCCTCAACACACATCTGCTTCA-----GTAACCTCTT	795
251	-----SerThrSerThrLeuThrGlyAsnHisPheThrIleValThrLeu	265
796	GACGTTGGATATCTTTGGCGGAGAAATTGGAATGAGGTTCACTTC	840
266	SerValCysHisPheGlyValGluLeuGlyGlyArgPheAsnPhe	280

RESULT 34

```

US-10-062-920-14'
; Sequence 14, Application US/10062920
; Publication NO. US20030096250A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/10/062,920
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US/09/660,587
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 14
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F
US-10-062-920-14

```

Alignment Scores:	
Pred. No.:	6.6e-57
Score:	605.00
Percent Similarity:	63.39%
Best Local Similarity:	45.08%
Query Match:	40.85%
DB:	9
Length:	280
Matches:	133
Conservative:	54
Mismatches:	78
Indels:	30
Gaps:	10

US-10-062-624-41 (1-840) x US-10-062-920-14 (1-280)

QY 1 ATGAATTATAAGAAATTCAGTAAGAACGCCGTTAATCTCATTAAATGTCAATCTTACCA 60  
||||| |||||| : : : : : ||||||| |||||||  
Db 1 MetAsnCysGlyLysPhePheIleThrThrLeuValSerLeuMetSerPheLeuPro 20





```

QY 508 TCATTGTTGTTAATCTGCTATGACATTAACAGCTGAAGAGTATCTTTTCGTACCATAT 567
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
171 SerLeuMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSerProTyr 190
QY 568 GCATGTGAGGTATAGGACGATCTTATCACTATTTTAAAGACCTCAATCAATAATT 627
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
191 IleCysAlaGlyValGlyThrAspLeuIleSerMetPheGluAlaIleAsnProLysIle 210
QY 628 GCTTACCAAGAAATAATAGTATTAGTTACCTCATCACACCAAGAGTCTCTGCAATTTAT 687
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
211 SerTyrGlnGlyLeuGlyLeuSerTyrSerIleSerProGluAlaSerValPheVal 230
QY 688 GGTGATACTACCATGGCTTATGGTAAATTAATTTGAGACATACCTGTAATACTCCT 747
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
231 GlyGlyHisPheHisLeysValIleGlyAsnGluPheArgAspIleProAlaMetIlePro 250
QY 748 GTAGTATTAAATGATGCTCCTCAACACCATCTGCTTCA-----GTAACCTCTT 795
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
251 -----SerThrSerThrLeuThrGlyAsnHisPheThrIleValThrLeu 265
QY 796 GACGTTGGATCTTTTGGCGGAGAAATTTGGAATGAGGTTTCACCTTC 840
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
266 SerValCysHisPheGlyValGluLeuGlyGlyArgPheAsnPhe 280

RESULT 37
US-10-059-964-12
; Sequence 12, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohashi, No. US20020120115A1
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; EARLIER FILING DATE: 2002-01-28
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 12
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-10-059-964-12

Alignment Scores:
Pred. No.: 6.6e-57 Length: 280
Score: 605.00 Matches: 133
Percent Similarity: 63.39% Conservative: 54
Best Local Similarity: 45.08% Mismatches: 78
Query Match: 40.85% Indels: 30
DB: 12 Gaps: 10

US-10-062-624-41 (1-840) x US-10-059-964-12 (1-280)
QY 1 ATGAATTATAAGAAAATCTAGTAAGACGCGTTAATCTCAATTAATGCTCAATCTTACCA 60
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 MetAsnCysLysLysPhePheIleThrThrLeuValSerLeuMetSerPheLeuPro 20
QY 61 TATCAGCTTTTGCAGATCTGTAGGTTCAAGAACTAATGATAACAAGAGGC---TTC 117
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
21 GlyIleSerPheSerAspAlaVal-----GlnAsnAspAsnValGlyGlyAsnPhe 37
QY 118 TACATTAGTCAAGTCAATCCCAAGTATATACACTTTAGAAAATTTCTGCTGCAAGAA 177
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
38 TyrIleSerGlyLysTyrValProSerValSerHisPheGlyValPheSerAlaLysGln 57
QY 178 ACTCCTATTAAATGAAACAAATTTCTCTCAAAAAAGTTTTCGGACTAAAGAAAGATGGT 237
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
58 -----GluArgAsnThrThrThrGlyValPheGlyLeuLysGlnAspTrp 72
QY 238 GAT-----ATAACAAAAAAGAC---GATTTTCAAGAGTAGTCTCCAGGCATTGAT 285

```

```

Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
73 AspGlySerThrIleSerLysAsnSerProGluAsnThrPheAsnValProAsnTyrSer 92
QY 286 TTT-----CAAAATAACTTAATATCAGGATTTTTCAGGAGTATTTGTTCTCTCTATGGAC 339
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
93 PheLysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaValGlyTyrLeuMetAsn 112
QY 340 GGACCAAGAATAGAACTTGAAGCTGCATATCAACAATTTTAAATCCAAAAAACACCGATAAC 399
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
113 GlyProArgIleGluLeuGluMetSerTyrGluThrPheAspValLysAsnGlnGlyAsn 132
QY 400 AATGATACTGATAATGTGTAATCTATATAACATTTTTCGATTATCTCGT-----447
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
133 AsnTyrLysAsnAspAla-----HisLysTyrTyrAlaLeuThrHisAsnSerGlyGly 150
QY 448 AAAGATCAATGGAAGATCAGCAATATGTAGTACTTAAATAATGACGGCATAACTTTTATG 507
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
151 LysLeuSerAsnAlaGlyAspLysPheValPheLeuLysAsnGluGlyLeuLeuAspIle 170
QY 508 TCATTGATGGTTAATCTGCTATGACATTCACACTGAAAGAGTATCTTTTCGTACCATAT 567
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
171 SerLeuMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSerProTyr 190
QY 568 GCATGTGAGGTATAGGACGATCTTATCACTATTTTAAAGACCTCAATCTCAATAATT 627
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
191 IleCysAlaGlyValGlyThrAspLeuIleSerMetPheGluAlaIleAsnProLysIle 210
QY 628 GCTTACCAAGAAATAATAGTATTAGTTACCTCATCACACCAAGAGTCTCTGCAATTTAT 687
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
211 SerTyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSerValPheVal 230
QY 688 GGTGATACTACCATGGCGTTTATTGTAATAAAATTTGAGAAGATACCTGTAATACTCCT 747
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
231 GlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProAlaMetIlePro 250
QY 748 GTAGTATTAAATGATGCTCCTCAACACCATCTGCTTCA-----GTAACCTCTT 795
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
251 -----SerThrSerThrLeuThrGlyAsnHisPheThrIleValThrLeu 265
QY 796 GACGTTGGATCTTTTGGCGGAGAAATTTGGAATGAGGTTTCACCTTC 840
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
266 SerValCysHisPheGlyValGluLeuGlyGlyArgPheAsnPhe 280

RESULT 38
US-10-284-986-18
; Sequence 18, Application US/10284986
; Publication No. US20030091588A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xu-Jie
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane
; TITLE OF INVENTION: Protein Multigene Family
; FILE REFERENCE: D6311D1
; CURRENT APPLICATION NUMBER: US/10/284,986
; PRIOR FILING DATE: 2002-10-31
; PRIOR FILING DATE: 09/846,808
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 18
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: P28-18 Outer Membrane Protein of
; OTHER INFORMATION: Ehrlichia chaffeensis
US-10-284-986-18

Alignment Scores:
Pred. No.: 1.09e-56 Length: 280
Score: 603.00 Matches: 133
Percent Similarity: 63.39% Conservative: 54
Best Local Similarity: 45.08% Mismatches: 78
Query Match: 40.72% Indels: 30

```

DB: 9 Gaps: 10

US-10-062-624-41 (1-840) x US-10-284-986-18 (1-280)

QY 1 ATGAATTATAAGAAATTCCTAGTAAGACGCGTTAAATCTCATTAAATGCTCAATCTTACCA 60  
| | | | | : : : : :  
Db 1 MetAenCysLysPhePheIleThrThrLeuValSerLeuMetSerPheLeuPro 20

QY 61 TATCAGTCTTTTGCAGATCTCTAGTTCACAGAACTAATGATAACAAAGAAGGC---TTC 117  
| | | | | : : : : :  
Db 21 GlyIleSerPheSerAspAlaVal-----GlnAsnAspValGlyGlyAsnPhe 37

QY 118 TACATTAGTCAAGTACAATCCAACTATATACACATTTAGAAAATTTCTCTGCTCAAGAA 177  
| | | | | : : : : :  
Db 38 TyrlleSerGlyLysTyrrValProSerValSerHisPheGlyValPheSerAlaLysGln 57

QY 178 ACTCCTATTAAATGGAACAAATTCCTCAGTAAAGATTTTCGGAAGTAAAGAAAGATGCT 237  
| | | | | : : : : :  
Db 58 -----GluArgAsnThrThrIleGlyValPheGlyLeuLysGlnAspTrp 72

QY 238 GAT-----ATAACAAAGAAAGAC---GATTTTACAGAGTAGCTCCAGGCTTGTAT 285  
| | | | | : : : : :  
Db 73 AspGlySerThrIleSerLysAsnSerProGluAsnThrPheAsnValProAsnTyrSer 92

QY 286 TTT-----CAAAATACTTAATATCAGGATTTTCAGGAAGTATTTGTTACTCTATGAC 339  
| | | | | : : : : :  
Db 93 PheLysTyrrGluAsnAsnProPheLeuGlyPheAlaGlyAlaValGlyTyrLeuMetAsn 112

QY 340 GGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAACACCCGATAAC 399  
| | | | | : : : : :  
Db 113 GlyProArgIleGluLeuGluMetSerTyrrGluThrPheAspValLysAsnGlnGlyAsn 132

QY 400 AATGATATCTGATAATGTTGTAATCTATATAACATTTTGCATTATCTCGT----- 447  
| | | | | : : : : :  
Db 133 AsnTyrrLysAsnAspAla-----HisLysTyrrTyrrAlaLeuThrHisAsnSerGlyGly 150

QY 448 AAAGATGCAATGGAAGATCAGCAATATGTAGTACTTAAATAATGACGCAATCTTTATG 507  
| | | | | : : : : :  
Db 151 LysLeuSerAsnAlaGlyAspLysPheValPheLeuLysAsnGluGlyLeuLeuAspTle 170

QY 508 TCATTGATGTTAATATCTTGTCTATGACATTCAGTGAAGGAGTATCTTCGTACCATAT 567  
| | | | | : : : : :  
Db 171 SerLeuMetLeuAsnAlaCysTyrrAspValIleSerGluGlyIleProPheSerProTyr 190

QY 568 GCATGTGAGGTATAGGACAGATCTTATCTACTATTTTAAAGACCTCAATCTAAAATTT 627  
| | | | | : : : : :  
Db 191 IleCysAlaGlyValGlyThrAspLeuIleSerMetPheGluAlaIleAsnProLysIle 210

QY 628 GCTTACCAAGAAATAAGTATTAGTTACCTATCACACCAGAGTCTCTGCTCATTTAT 687  
| | | | | : : : : :  
Db 211 SerTyrrGlnGlyLysLeuGlyLeuSerTyrrSerIleSerProGluAlaSerValPheVal 230

QY 748 GTAGTATTAAATGATGCTCCTCAACACATCTGCTTCA-----GTAATCTCTT 795  
| | | | | : : : : :  
Db 251 -----SerThrSerThrLeuThrGlyAsnHisPheThrIleValThrLeu 265

QY 796 GACGTTGGATCTTTGGCGGAAATTTGGAATGAGGTTCCACCTTC 840  
| | | | | : : : : :  
Db 266 SerValCysHisPheGlyValGluLeuGlyGlyArgPheAsnPhe 280

## RESULT 39

US-09-846-808-18  
; Sequence 18, Application US/09846808  
; Patent No. US20020064531A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: Yu, Xu-Jie  
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane  
; TITLE OF INVENTION: Protein Multigene Family

; FILE REFERENCE: D6311  
; CURRENT APPLICATION NUMBER: US/09/846,808  
; CURRENT FILING DATE: 2001-05-01  
; PRIOR APPLICATION NUMBER: 60/201,035  
; PRIOR FILING DATE: 2000-05-01  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 18  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: P28-18 Outer Membrane Protein of  
; OTHER INFORMATION: Ehrlichia chaffeensis  
US-09-846-808-18

Alignment Scores:  
Pred. No.: 1,09e-56 Length: 280  
Score: 603.00 Matches: 133  
Percent Similarity: 63.39% Conservative: 54  
Best Local Similarity: 45.08% Mismatches: 78  
Query Match: 40.72% Indels: 30  
DB: 10 Gaps: 10

US-10-062-624-41 (1-840) x US-09-846-808-18 (1-280)

QY 1 ATGAATTATAAGAAATTCCTAGTAAGACGCGTTAAATCTCATTAAATGCTCAATCTTACCA 60  
| | | | | : : : : :  
Db 1 MetAenCysLysPhePheIleThrThrLeuValSerLeuMetSerPheLeuPro 20

QY 61 TATCAGTCTTTTGCAGATCTCTAGTTCACAGAACTAATGATAACAAAGAAGGC---TTC 117  
| | | | | : : : : :  
Db 21 GlyIleSerPheSerAspAlaVal-----GlnAsnAspValGlyGlyAsnPhe 37

QY 118 TACATTAGTCAAGTACAATCCAACTATATACACATTTAGAAAATTTCTCTGCTCAAGAA 177  
| | | | | : : : : :  
Db 38 TyrlleSerGlyLysTyrrValProSerValSerHisPheGlyValPheSerAlaLysGln 57

QY 178 ACTCCTATTAAATGGAACAAATTCCTCAGTAAAGATTTTCGGAAGTAAAGAAAGATGCT 237  
| | | | | : : : : :  
Db 58 -----GluArgAsnThrThrIleGlyValPheGlyLeuLysGlnAspTrp 72

QY 238 GAT-----ATAACAAAGAAAGAC---GATTTTACAGAGTAGCTCCAGGCTTGTAT 285  
| | | | | : : : : :  
Db 73 AspGlySerThrIleSerLysAsnSerProGluAsnThrPheAsnValProAsnTyrSer 92

QY 286 TTT-----CAAAATACTTAATATCAGGATTTTCAGGAAGTATTTGTTACTCTATGAC 339  
| | | | | : : : : :  
Db 93 PheLysTyrrGluAsnAsnProPheLeuGlyPheAlaGlyAlaValGlyTyrLeuMetAsn 112

QY 340 GGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAACACCCGATAAC 399  
| | | | | : : : : :  
Db 113 GlyProArgIleGluLeuGluMetSerTyrrGluThrPheAspValLysAsnGlnGlyAsn 132

QY 400 AATGATATCTGATAATGTTGTAATCTATATAACATTTTGCATTATCTCGT----- 447  
| | | | | : : : : :  
Db 133 AsnTyrrLysAsnAspAla-----HisLysTyrrTyrrAlaLeuThrHisAsnSerGlyGly 150

QY 448 AAAGATGCAATGGAAGATCAGCAATATGTAGTACTTAAATAATGACGCAATCTTTATG 507  
| | | | | : : : : :  
Db 151 LysLeuSerAsnAlaGlyAspLysPheValPheLeuLysAsnGluGlyLeuLeuAspTle 170

QY 508 TCATTGATGTTAATATCTTGTCTATGACATTCAGTGAAGGAGTATCTTCGTACCATAT 567  
| | | | | : : : : :  
Db 171 SerLeuMetLeuAsnAlaCysTyrrAspValIleSerGluGlyIleProPheSerProTyr 190

QY 568 GCATGTGAGGTATAGGACAGATCTTATCTACTATTTTAAAGACCTCAATCTAAAATTT 627  
| | | | | : : : : :  
Db 191 IleCysAlaGlyValGlyThrAspLeuIleSerMetPheGluAlaIleAsnProLysIle 210

QY 628 GCTTACCAAGAAATAAGTATTAGTTACCTATCACACCAGAGTCTCTGCTCATTTAT 687  
| | | | | : : : : :  
Db 211 SerTyrrGlnGlyLysLeuGlyLeuSerTyrrSerIleSerProGluAlaSerValPheVal 230

```
QY 688 GGTGATACCTACCATGGCTTATTCGTAATAAATTTGAGAAGATACCTGTAATAACTCCT 747
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
231 GlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProAlaMetIlePro 250
QY 748 GTAGTATTAAATGATGCTCCTCAACACCATCTGCTTCA-----GTAACCTCT 795
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251 -----SerThrSerThrLeuThrGlyAsnHisPheThrIleValThrLeu 265
QY 796 GACGTTGATACCTTTGGCGGAGAAATTTGGAATGAGTTTCACCTTC 840
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
266 SerValCysHisPheGlyValGluLeuGlyGlyArgPheAsnPhe 280

RESULT 40
US-10-062-994-17
; Sequence 17, Application US/10062994
; Patent No. US20010001661A1
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Bowie, Michael V.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Mahan, Suman M.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Rutangirwa, Fred R.
; APPLICANT: Moreland, Annie L.
; APPLICANT: Simbi, Bigboy H.
; APPLICANT: Whitcure, William M.
; APPLICANT: Alleman, Arthur R.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and
; FILE REFERENCE: UF-167XC3
; CURRENT APPLICATION NUMBER: US/10/062,994
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 09/533,662
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 09/337,827
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 08/953,326
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 08/733,230
; PRIOR FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 280
; TYPE: PRN
; ORGANISM: Ehrlichia chaffeensis
US-10-062-994-17

Alignment Scores:
Pred. No.: 1.09e-56 Length: 280
Score: 603.00 Matches: 133
Percent Similarity: 63.39% Conservative: 54
Best Local Similarity: 45.08% Mismatches: 78
Query Match: 40.72% Indels: 30
DB: 12 Gaps: 10

US-10-062-624-41 (1-840) x US-10-062-994-17 (1-280)
QY 1 ATGAATTATAAGAAATTCATAGAACGCGTAAATCTCATTAATGTCATCTTACCA 60
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MetAsnCysLysLysPhePheIleThrThrLeuValSerLeuMetSerPheLeuPro 20
QY 61 TATCAGTCTTTTCAGATCCCTGAGTTTCAAGAACTAATGATAACAAAGAGGC---TTC 117
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
21 GlyIleSerPheSerAspAlaVal-----GlnAsnAspAsnValGlyGlyAsnPhe 37
QY 118 TACATTAGTCAAGTACAAATCCAGTATATACACTTTAGAAATTTCTGCTGAAGAA 177
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
38 TyrIleSerGlyLysTyValProSerValSerHisPheGlyValPheSerAlaLysGln 57
QY 178 ACTCCTATTATGACAAATCTCTCACTAAAGAGTTTTCGGACTAAAGAAAGATGCT 237
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
58 -----GluArgAsnThrThrIleGlyValPheGlyLeuLysGlnAspTrp 72
```

```
QY 238 GAT-----ATAACAAAAAAGAC---GATTTTCAAGAGTAGTCTCCAGGCATTGAT 285
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
73 AspGlySerThrIleSerLysAsnSerProGluAsnThrPheAsnValProAsnTyrSer 92
QY 286 TTT-----CAAAATAACTTAATATCAGGATTTTCAGGAAGTATTTGTTACTCTATGAC 339
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
93 PheLysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaValGlyTyrLeuMetAsn 112
QY 340 GGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAAACACCGATAAC 399
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
113 GlyProArgIleGluLeuGluMetSerTyrGluThrPheAspValLysAsnGlnGlyAsn 132
QY 400 AATGATACTGATAATGTTGTAATCTATAAATTTTGCATTATCTCTCGT----- 447
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
133 AsnTyrLysAsnAspAla-----HisLysTyrTyrAlaLeuThrHisAsnSerGlyGly 150
QY 448 AAAGATGCAATGGAAGATCAGCAATATGTACTTCTATAAATGAGCGCATATCTTTATG 507
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 LysLeuSerAsnAlaGlyAspLysPheValPheLeuLysAsnGluGlyLeuLeuAspIle 170
QY 508 TCATTGATGGTTAATACTTGTATGACATTACAGCTGAAGAGGATATCTTTCGTACCATAT 567
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
171 SerLeuMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSerProTyr 190
QY 568 GCATGTGCAGGTATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAAAATT 627
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
191 IleCysAlaGlyValGlyThrAspLeuIleSerMetPheGluAlaIleAsnProLysIle 210
QY 628 GCTTACCAAGAAAAATAGTATTTAGTTACCTATACACCAAGAGTCTCTGCATTTAT 687
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
211 SerTyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSerValPheVal 230
QY 688 GGTGATACTACCATCGCTTATTGGTAATAAATTTGAGAAGATACCTGTAATAACTCCT 747
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
231 GlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProAlaMetIlePro 250
QY 748 GTAGTATTAAATGATGCTCCTCAACACCATCTCTCTTCA-----GTAACCTCT 795
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251 -----SerThrSerThrLeuThrGlyAsnHisPheThrIleValThrLeu 265
QY 796 GACGTTGATACCTTTGGCGGAGAAATTTGGAATGAGTTTCACCTTC 840
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
266 SerValCysHisPheGlyValGluLeuGlyGlyArgPheAsnPhe 280
```

Search completed: July 8, 2003, 11:04:30  
Job time : 67.5 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 8, 2003, 10:41:46 ; Search time 42.5 Seconds  
(without alignments)  
3800.137 Million cell updates/sec

Title: US-10-062-624-41  
Perfect score: 1481  
Sequence: 1 atgaattataagaataattct.....ttgggaatgaggttcaccttc 840

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Deloop 6.0 , Deloext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool/US10062624/runat\_08072003\_093520\_2027/app.query.fasta\_1.1031  
-DB=pir\_73 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0 -I-LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USR=US10062624 @CGN 1 1 62 @runat\_08072003\_093520\_2027 -NCPU=3  
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOOP=6 -DELOEXT=7

Database : PIR\_73.\*

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	629.5	42.5	276	2	28k surface antige
2	620	41.9	286	2	28k surface antige
3	603	40.7	280	2	28k surface antige
4	577	39.0	278	2	28k surface antige
5	565	38.1	284	2	major antigenic pr
6	515.5	34.8	287	2	28k surface antige
7	386.5	26.1	282	2	major surface prot
8	227.5	15.4	133	2	28k surface antige
9	119.5	8.1	264	2	heat resistant agg
10	107.5	7.3	239	2	probable outer mem
11	104.5	7.1	738	2	hemolysin A precu
12	103	7.2	1613	2	hypothetical prote
13	102	6.9	1176	2	DNA-directed RNA p
14	101.5	6.9	415	2	3-oxoacyl-l-acyl-ca

15	100	6.8	1582	2	AC1153	adhesin homolog lm
16	98	6.6	3890	2	C89921	hypothetical prote
17	97.5	6.6	274	2	C83295	heat resistant agg
18	97.5	6.6	1174	2	S28976	DNA-directed RNA p
19	97.5	6.6	1271	2	A45555	glutamate rich pro
20	97.5	6.6	1310	2	AD1380	glycoelidase homolo
21	95	6.6	665	2	D90092	hypothetical prote
22	95	6.4	1004	2	B25039	outer cell wall pr
23	95	6.4	1578	2	AD1512	peptidoglycan boun
24	94.5	6.4	474	2	F97264	6-Phospho-Beta-D-G
25	93.5	6.3	658	2	E95111	endo-beta-N-acetyl
26	93	6.3	180	2	F71639	hypothetical prote
27	93	6.3	608	2	H64473	hypothetical prote
28	93	6.3	745	2	T51370	hypothetical prote
29	93	6.3	1348	2	B23496	TYB protein - yeas
30	93	6.3	1598	2	S69967	TYB protein - yeas
31	93	6.3	1770	2	S69953	TYB protein - yeas
32	93	6.3	1770	2	S70233	TYB protein - yeas
33	93	6.3	1770	2	S70230	TYB protein - yeas
34	93	6.3	1770	2	S69966	TYB protein - yeas
35	93	6.3	1770	2	S69950	TYB protein - yeas
36	93	6.3	1770	2	S58651	TYB protein - yeas
37	93	6.3	1771	2	S53592	TYB protein - yeas
38	93	6.3	1810	2	S69973	hypothetical prote
39	92.5	6.2	219	2	C97440	conserved hypoteth
40	92.5	6.2	219	2	AF2658	opacity protein op
41	92.5	6.2	257	2	S16614	prophage pi2 prote
42	92.5	6.2	401	2	F86754	probable membrane
43	92.5	6.4	532	2	S46831	probable electron
44	92.5	6.2	585	2	C69336	endo-beta-N-acetyl
45	91.5	6.2	721	2	C97980	

#### ALIGNMENTS

##### RESULT 1

JE0218  
28k surface antigen 5 - Ehrlichia chaffensis  
N:Alternate names: MAP1  
C:Species: Ehrlichia chaffensis  
C:Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 17-Mar-1999  
C:Accession: JE0218  
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrige, M.J.; Allenan, A.R.  
Biochem. Biophys. Res. Commun. 247, 636-643, 1998  
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe  
A:Reference number: JE0216; MUID:98321180; PMID:9647746  
A:Accession: JE0218  
A:Molecule type: DNA  
A:Residues: 1-276 <RED>  
A:Cross-references: GB:AF062761

Alignment Scores:  
Pred. No.: 3 66e-49 Length: 276  
Score: 629.50 Matches: 137  
Percent Similarity: 62.24% Conservatives: 41  
Best Local Similarity: 47.90% Mismatches: 87  
Query Match: 42.51% Indels: 21  
DB: 2 Gaps: 6

US-10-062-624-41 (1-840) x JE0218 (1-276)

Qy	1	ATGAATTATAGAAATTTCTAGTAAGAGCGGTTAATCTCAATTAATCTCAATCTTACCA	60
Db	1	MetAsnTyrLysValPheIleThrSerAlaLeuIleSerLeuIleSerSerLeuPro	20
Qy	61	TATCAGCTTTTCAGATCTCTAGGTTCAAGAACTAATCATATACAAGAGGCTTCTAC	120
Db	21	GlyValSerPheSerAspProAlaGlySerGlyIleAsnGlyAsn-----PheTyr	37
Qy	121	ATTAGTCCAAGTACATCAAGTATATACATATACACATTAGAAAATTTCTCTGTGAAGAACT	180
Db	38	IleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu---	56

QY 181 CCTATTATGGAACAAATTTCTCTCACTAAAGAGTTTTCGGACTAAAGAGAGATGGTGAT 240  
 Db : : : : :  
 QY 57 -----GluArgAsnThrThrValGlyValPheGlyLeuGlyGlnAsnTrpAsp 72  
 Db : : : : :  
 QY 241 -----ATAACAAAAGACGATTTTACAGAGTAGTCCAGGCATT 282  
 Db : : : : :  
 QY 73 GlySerAlaIleSerAsnSerProAsnAspValPheThrValSerAsnTyrSerPhe 92  
 Db : : : : :  
 QY 283 GATTTTCAAAATACTTAATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGACCGGA 342  
 Db : : : : :  
 QY 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetAspGly 112  
 Db : : : : :  
 QY 343 CCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAACACCGATACAAT 402  
 Db : : : : :  
 QY 113 ProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn 132  
 Db : : : : :  
 QY 403 GATACCTGATAATGGTGAATCTATATAACATTTTGCATTATCTCGTAAAGATGCAATGGA 462  
 Db : : : : :  
 QY 133 -----TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerHisAsnSerAlaAlaAsp 150  
 Db : : : : :  
 QY 463 -----GATCAGCAATATGTAGTACTTAATAATGACGGCATAACTTTTATGTCA 510  
 Db : : : : :  
 QY 151 MetSerSerAlaSerAsnAsnPheValPheLeuLysAsnGluGlyLeuLeuAspIleSer 170  
 Db : : : : :  
 QY 511 TTGATGGTTAATACTTGTATGACATTACAGCTGAAGGAGTATCTTTCTGTACCATATGCA 570  
 Db : : : : :  
 QY 171 PheMetLeuAsnAlaCysTyrAspValValGlyGluGlyIleProPheSerProTyrIle 190  
 Db : : : : :  
 QY 571 TGTGAGTATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTTGGT 630  
 Db : : : : :  
 QY 191 CysAlaGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSer 210  
 Db : : : : :  
 QY 631 TACCAGAAAATAGTATTAGTTACCTATACACAGAGTCTCTGCATTTATTTGGT 690  
 Db : : : : :  
 QY 211 TyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSerValPheIleGly 230  
 Db : : : : :  
 QY 691 GGATACTACCATGGGTTATTGGTAAATAATTTGAGAAGACCTGTAACTCTCTGTA 750  
 Db : : : : :  
 QY 231 GlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrIleIleProThr 250  
 Db : : : : :  
 QY 751 GTATTAAATGATGCTCTCTCAA---ACCACATCTGCTTCAGTAACCTCTTCAGCTGGATAC 807  
 Db : : : : :  
 QY 251 GlySerThrLeuAlaGlyLysGlyAsnTyrProAlaIleValIleLeuAspValCysHis 270  
 Db : : : : :  
 QY 808 TTGCGCGAGAAATGGA 825  
 Db : : : : :  
 QY 271 PheGlyIleGluMetGly 276  
 Db : : : : :  
 RESULT 2  
 JE0219  
 28k surface antigen 2 - Ehrlichia chaffensis  
 N:Alternate names: MAP1  
 C:Species: Ehrlichia chaffensis  
 C>Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 17-Mar-1999  
 C:Accession: JE0219  
 R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.  
 Biochem. Biophys. Res. Commun. 247, 636-643, 1998  
 A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe  
 A:Reference number: JE0216; MUID:98321180; PMID:9647746  
 A:Accession: JE0219  
 A:Molecule type: DNA  
 A:Residues: 1-286 <RED>  
 A:Cross-references: GB:AF062761  
 Alignment Scores:  
 Pred. No.: 2.68e-48 Length: 286  
 Score: 620.00 Matches: 139  
 Percent Similarity: 41.56% Conservative: 42  
 Best Local Similarity: 47.28% Mismatches: 89  
 Query Match: 41.86% Indels: 24  
 DB: 2 Gaps: 7  
 US-10-062-624-41 (1-840) x JE0219 (1-286)

QY 1 ATGAATTATAGRAAAATCTAGTAGAAGCGCGTTAATCTCATTAATTAATCTCAATCTTACCA 60  
 Db : : : : :  
 QY 1 MetAenCysGluLysPhePheIleThrAlaLeuThrLeuLeuMetSerPheLeuPro 20  
 Db : : : : :  
 QY 61 TATCAGTCTTTTTCAGAGATCTCTGTAGGTTCAAGAACTAATGATATAACAAAGAGGC---TTC 117  
 Db : : : : :  
 QY 21 GlyIleSerLeuSerAspProVal-----GlnAspAspAsnIleSerGlyAsnPhe 37  
 Db : : : : :  
 QY 118 TACATTAGTGCAGAGTACAATCCAGTATATCAACATTTAGAAAAATTTCTCTGTGGAAGAA 177  
 Db : : : : :  
 QY 38 TyrIleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu 57  
 Db : : : : :  
 QY 178 ACTCTTATTATGAACAAATTTCTCTCACTAAAGATTTTTCGACATAAAGAGATGGT 237  
 Db : : : : :  
 QY 58 -----GluArgAsnThrThrValGlyValPheGlyIleGluGlnAspTrp 72  
 Db : : : : :  
 QY 238 GAT-----ATAACAAAAGACGATTTTACAAGAGTAGTCCAGGC 279  
 Db : : : : :  
 QY 73 AspArgCysValIleSerArgThrThrLeuSerAspIlePheThrValProAsnTyrSer 92  
 Db : : : : :  
 QY 280 ATTGATTTTCAAAATACTTAATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGAC 339  
 Db : : : : :  
 QY 93 PheLysTyrGluAsnAsnLeuPheSerGlyPheAlaGlyAlaIleGlyTyrSerMetAsp 112  
 Db : : : : :  
 QY 340 GGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAACACCGATAC 399  
 Db : : : : :  
 QY 113 GlyProArgIleGluLeuGluValSerTyrGluAlaPheAspValLysAsnGlnGlyAsn 132  
 Db : : : : :  
 QY 400 AATGATACCTGATAATGGTGAATCTATATAACATTTTGCATTATCTCGTAAAGATGCAATG 459  
 Db : : : : :  
 QY 133 Asn-----TyrLysAsnGluAlaHisArgTyrTyrAlaLeuSerHisLeuLeuGlyThr 150  
 Db : : : : :  
 QY 460 GAAGATCAG-----CAATATCTAGTACTTAAATAATGACGGATAACT 501  
 Db : : : : :  
 QY 151 GluThrGlnIleAspGlyAlaGlySerAlaSerValPheLeuIleAsnGluGlyLeu 170  
 Db : : : : :  
 QY 502 TTTATGTCATGTAGTGTATATCTGCTGTATGACATTACAGCTGAAGGAGTATCTTCGTA 561  
 Db : : : : :  
 QY 171 AspLysSerPheMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSer 190  
 Db : : : : :  
 QY 562 CCATATGATGTGAGGATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTA 621  
 Db : : : : :  
 QY 191 ProTyrIleCysAlaGlyIleGlyIleAspLeuValSerMetPheGluAlaIleAsnPro 210  
 Db : : : : :  
 QY 622 AAATTTGCTTACCAAGGAAAATAGGTATTAGTTACCTATCACACCGAAGTCTCTGCA 681  
 Db : : : : :  
 QY 211 LysIleSerTyrGlnGlyLysLeuGlyLeuSerTyrProIleSerProGluAlaSerVal 230  
 Db : : : : :  
 QY 682 TTTATGCTGGATACTACCATGGCTTATTGGTAAATAATTTGAGAAGATACCTGTAATA 741  
 Db : : : : :  
 QY 231 PheIleGlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrMet 250  
 Db : : : : :  
 QY 742 ACTCTGTAGTATTAAATGATGCTCTCTCAA---ACCACATCTGCTTCAGTAACCTCTGAC 798  
 Db : : : : :  
 QY 251 IleProSerGluSerAlaLeuAlaGlyLysGlyAsnTyrProAlaIleValThrIleAsp 270  
 Db : : : : :  
 QY 799 GTTGGATCTTTTTCGGAGAAATTTGGAATGAGTTACCTTCACCTTC 840  
 Db : : : : :  
 QY 271 ValPheTyrPheGlyIleGluLeuGlyGlyArgPheAsnPhe 284  
 Db : : : : :  
 RESULT 3  
 JE0217  
 28k surface antigen 4 - Ehrlichia chaffensis  
 N:Alternate names: MAP1  
 C:Species: Ehrlichia chaffensis  
 C>Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 17-Mar-1999  
 C:Accession: JE0217  
 R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.  
 Biochem. Biophys. Res. Commun. 247, 636-643, 1998  
 A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe  
 A:Reference number: JE0216; MUID:98321180; PMID:9647746  
 A:Accession: JE0217

A:Molecule type: DNA  
A:Residues: 1-280 <RED>  
A:Cross-references: GB:AF062761

Alignment Scores:  
Pred. No.: 9.38e-47 Length: 280  
Score: 603.00 Matches: 133  
Percent Similarity: 63.59% Conservative: 54  
Best Local Similarity: 45.08% Mismatches: 78  
Query Match: 40.72% Indels: 30  
DB: 2 Gaps: 10

US-10-062-624-41 (1-840) x JE0217 (1-280)

```
QY 1 ATGAATTATAGAAAAATTCAGTAAAGCGCGTTAATCTCAATTAATCTCAATCTTACCA 60
DB 1 MetAsnCysLysLysPhePheIleThrThrLeuValSerLeuMetSerPheLeuPro 20

QY 61 TATCAGTCTTTGAGATCCTGTAGTTCAGAACTAATGATACAAAGAGGC---TTC 117
DB 21 GlyIleSerPheSerAspAlaVal-----GlnAsnAspAsnValGlyGlyAsnPhe 37

QY 118 TACATTAGTCAAAAGTACAAATCCAAAGTATATCACACTTTAGAAAAATCTCTGCTGAAGAA 177
DB 38 TyzIleSerGlyLysTyrrValProSerValSerHisPheGlyValPheSerAlaLysGln 57

QY 178 ACTCCTATTAAATGAACAAATTTCTCTCACTAAAAAGTTTTCGAGCTAAAGAAAGATGGT 237
DB 58 -----GluArgAsnThrThrIleGlyValPheGlyLeuLysGlnAspTyr 72

QY 238 GAT-----ATAACAAAAAGAC---GATTTTACAGAGTAGTCTCCGGCATTTGAT 285
DB 73 AspGlySerThrIleSerLysAsnSerProGluAsnThrPheAsnValProAsnTyrSer 92

QY 286 TTT-----CAAAATAACTTAATATCAGGATTTTCAGAGACTATTGGTTACTCTCATGAC 339
DB 93 PheLysTyrGluAsnAsnPropheLeuGlyPheAlaGlyValGlyTyrLeuMetAsn 112

QY 340 GGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTTAAATCCAAAAACACCGATAAC 399
DB 113 GlyProArgIleGluLeuGluMetSerTyrGluThrPheAspValLysAsnGlnGlyAsn 132

QY 400 AATGATACTGATGTAATGTAATCTATCAATAATTTTGCATTATCTCGT----- 447
DB 133 AsnTyrLysAsnAspAla-----HisLysTyrTyrAlaLeuThrHisAsnSerGlyGly 150

QY 448 AAGATCAATGGAAGATCAGCAATATGTAGTACTTAAATAGCGGCAATCTTTATG 507
DB 151 LysLeuSerAsnAlaGlyAspLysPheValPheLeuLysAsnGluGlyLeuLeuAspIle 170

QY 508 TCATTGATGGTTAATCTATGCTATGACATTACAGCTGAAGAGGATPCTCTTCGTACCATAT 567
DB 171 SerLeuMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSerProTyr 190

QY 568 GCATGTCAGGTATAGAGCGATCTTATCATCTATTTTAAAGACCTCAATCTAAATTT 627
DB 191 IleCysAlaGlyValGlyThrAspLeuIleSerMetPheGluAlaIleAsnProLysIle 210

QY 628 GCTTACCAGGAAATAGGTATTAGTTACCTATCACACAGAGTCTCTGCAATTTAT 687
DB 211 SerTyrGlnGlyLysLeuSerTyrSerIleSerProGluAlaSerValPheVal 230

QY 688 GGTGATACCTACCATGGCGTTATTGGTAATAATTTTGAAGAATACCTGTAATACTCCT 747
DB 231 GlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProAlaMetIlePro 250

QY 748 GTAGTATTAAATGATGCTCTCAACACCATCTCTCTCA-----GTAACCTCT 795
DB 251 -----SerThrSerThrLeuThrGlyAsnHisPheThrIleValThrLeu 265

QY 796 GACGTTGATCTTTGGCGGAGAAATTTGGATGAGGTTCCACCTTC 840
DB 266 SerValCysHisPheGlyValGluLeuGlyGlyArgPheAsnPhe 280
```

RESULT 4  
JE0216

28k surface antigen 3 - Ehrlichia chaffensis  
N:Alternate names: MAP1  
C:Species: Ehrlichia chaffensis  
C>Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 17-Mar-1999  
C/Accession: JE0216  
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; BurrIDGE, M.J.; Alleman, A.R.  
Biochem. Biophys. Res. Commun. 247, 636-643, 1998  
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe  
A:Reference number: JE0216; MUID:98321180; PMID:9647746  
A:Accession: JE0216

A:Molecule type: DNA  
A:Residues: 1-278 <RED>  
A:Cross-references: GB:AF062761

Alignment Scores:  
Pred. No.: 2.16e-44 Length: 278  
Score: 577.00 Matches: 126  
Percent Similarity: 60.55% Conservative: 49  
Best Local Similarity: 43.60% Mismatches: 94  
Query Match: 38.96% Indels: 20  
DB: 2 Gaps: 7

US-10-062-624-41 (1-840) x JE0216 (1-278)

```
QY 1 ATGAATTATAGAAAAATTCAGTAAAGCGCGTTAATCTCAATTAATCTCAATCTTACCA 60
DB 1 MetAsnCysLysLysPhePheIleThrThrAlaLeuValSerLeuMetSerPheLeuPro 20

QY 61 TATCAGTCTTTGAGATCCTGTAGTTCAGAACTAATGATACAAAGAGGC---TTC 117
DB 21 GlyIleSerPheSerAspProVal-----GlnGlyAsnIleSerGlyAsnPhe 37

QY 118 TACATTAGTCAAAAGTACAAATCCAAAGTATATCACACTTTAGAAAAATCTCTGCTGAAGAA 177
DB 38 TyzValSerGlyLysTyrMetProSerAlaSerHisPheGlyMetPheSerAlaLysGlu 57

QY 178 ACTCCTATTAAATGAACAAATTTCTCTCACTAAAAAGTTTTCGAGCTAAAGAAAGAT--- 234
DB 58 -----GluLysAsnProThrValAlaLeuTyrGlyLeuLysGlnAspTyr 72

QY 235 -----GGTGATATACAAAAAGACGATTTTACAAAGAGTAGCTCCAGGCATT 282
DB 73 GluGlyIleSerSerSerHisAsnAspAsnHisPheAsnAsnLysGlyTyrSerPhe 92

QY 283 GATTTTCAAAATTAATTAATCAGGATTTTCAGGAAGTATTGGTTACTCTATGCGCGGA 342
DB 93 LysTyrGluAsnAsnPropheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGlyGly 112

QY 343 CCAAGATAGAACTTGAAGCTGCATATCAACAATTTTAAATCCAAAAACACCGATAACAAT 402
DB 113 ProArgValGluPheGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsn 132

QY 403 GATACGATAAATGGTGAATACTATTAACATTTTGCATTTCTCGTAAAGAT----- 453
DB 133 TyrLysAsnAspAla-----HisArgTyrCysAlaLeuGlyGlnGlnAspAsnSerGly 150

QY 454 GCATGGAAGATCAGCAATATGTAGTACTTAAAAATGACGCGCATAACTTTATGTCATTG 513
DB 151 IleProLysThrSerLysTyrValLeuLeuLysSerGlyLeuLeuAspIleSerPhe 170

QY 514 ATGTTTAATCTGCTATGACATACAGCTGGAAGGAGTATCTTTTCGTACCATGTCATGT 573
DB 171 MetLeuAsnAlaCysTyrAspIleIleAsnGluSerIleProLeuSerProTyrIleCys 190

QY 574 GCAGGTATAGGACAGATCTTATCATTATTTTAAAGACCTCAATCTTAAATTTGCTTAC 633
DB 191 AlaGlyValGlyThrAspLeuIleSerMetPheGluAlaThrAsnProLysIleSerTyr 210

QY 634 CAAGGAAAAATAGGTATTAGTTACCTTATCACCCTATCACACGAGAGTCTCTGCAATTATTGGTGA 693
```







A;Cross-references: GB:AF062762; NID:g3327964; PIDN:AAC26722.1; PID:g3327966

## Alignment Scores:

Pred. No.: 1.15e-12 Length: 133  
Score: 227.50 Matches: 58  
Percent Similarity: 55.71% Conservative: 20  
Best Local Similarity: 41.43% Mismatches: 49  
Query Match: 15.36% Indels: 13  
DB: 2 Gaps: 4

US-10-062-624-41 (1-840) x JB0221 (1-133)

QY 1 ATGATTATAGAAATCTAGTAGAAGCGGTTAATCTCATTAATCTCAATCTTACCA 60  
DB 1 MetAsnCysLysValPheThrIleSerAlaLeuIleSerSerIlePheLeuPro 20  
QY 61 TATCAGTCTTTGAGAGTCTCTGAGTCTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120  
DB 21 AsnValSerTyrSerAsnProValTyrGly-----AsnSerMetTyrGlyAsnPheTyr 38  
QY 121 ATTAGTCAAGTACAAATCCAGTATATACACATTTAGAAAATCTCTGCTGAGAAACT 180  
DB 39 IleSerGlyLysTyrMetProSerValProHisPheGlyIlePheSerAlaGluGlu--- 57  
QY 181 CCTATTATGGAACAAATCTCTCACTAAAGTTTTCGGACTAAAGAAAGAT----- 234  
DB 58 -----GluLysLysLysThrValValTyrGlyLeuLysGluAsnTrpAla 73  
QY 235 GGTGATATATA-----ACAAAAGAGCATTTTACAAAGTAGTAGCTCCAGGCATT 282  
DB 74 GlyAspAlaIleSerSerGlnSerProAspAsnPheThrIleArgAsnTyrSerPhe 93  
QY 283 GATTTCAAAATACTAATATATCAGATTTTCAGGAAGTATGGTTACTTATGAGACGGA 342  
DB 94 LysTyrAlaSerAsnLysPheLeuGlyPheAlaValAlaIleGlyTyrSerIleGlySer 113  
QY 343 CCAGAATAGACTTGAAGCTGCATATCAACATTTTATCCAAAACACCGATACAAAT 402  
DB 114 ProArgIleGluValGluMetSerTyrGluAlaPheAspValLysGlnGlyAsnAsn 133

## RESULT 9

154668  
heat resistant agglutinin 1 - Escherichia coli  
C;Species: Escherichia coli  
C;Date: 07-Jun-1996 #sequence revision 07-Jun-1996 #text\_change 08-Oct-1999  
R;Accession: 154668; 169133; 169134  
R;Lutwyche, P.; Rupps, R.; Cavanagh, J.; Warren, R.A.; Brooks, D.E.  
Infect. Immun. 62, 5020-5026, 1994  
A;Title: Cloning, sequencing, and viscometric adhesion analysis of heat-resistant agglutinin  
A;Reference number: 154668; MUID:95012721; PMID:7927783  
A;Accession: 154668  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-264 <RES>  
A;Cross-references: EMBL:U07174; NID:g463910; PIDN:AAC13751.1; PID:g463911  
A;Accession: 169133  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 15-264 <RES>  
A;Cross-references: EMBL:U07174; NID:g463910; PIDN:AAC13752.1; PID:g463912  
A;Accession: 169134  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 18-264 <RES>  
A;Cross-references: EMBL:U07174; NID:g463910; PIDN:AAC13753.1; PID:g463913

## Alignment Scores:

Pred. No.: 0.00805 Length: 264  
Score: 119.50 Matches: 59  
Percent Similarity: 38.91% Conservative: 41  
Best Local Similarity: 22.96% Mismatches: 66  
Query Match: 8.07% Indels: 91  
DB: 2 Gaps: 14

US-10-062-624-41 (1-840) x 154668 (1-264)

QY 13 AAAATCTAGTAAGAAGCGGTTAATCTCATTAATGTCATATCTTACCATATCAGTCTTTT 72  
DB 20 LysValIleAlaValSerAlaLeu---AlaMetAlaGlyMetPheSerThrGlnAlaLeu 38  
QY 73 GCAGATCTGTAGGTTCAAGNACTAATGATAACAAAGAGGCTTCTACATTTAGTGCAAG 132  
DB 39 AlaAsp-----GluSerLysThrGlyPheTyrValThrGlyLys 51  
QY 133 TACAATCCAAGTATATACACTTTTAGAAAATCTCTCTGCTGAAGAAATCTCTTATTATGGA 192  
DB 52 AlaGlyAlaSerValMetSerLeu-----AlaAspGlnAArgPheLeuSerGly 67  
QY 193 ACAAAATCTCTCACTAAAAAGTTTTCGGACTAAAGAAAGATGTGTATATAACAAAAA 252  
DB 68 AsnGlyGluGluThrSerLysTyr-----LysGlyGlyAsp----- 79  
QY 253 GACGATTTTACAGAGTAGCTCCAGGCATTGATTTTCAAAAATACTTAATATCAGGATTT 312  
DB 80 -----GlyHisAsp-----ThrValPhe 85  
QY 313 TCAGAAAGTATGTTTACTCTATGGAC-----GGACCAAGAATA 351  
DB 86 SerGlyGlyIleAlaAlaGlyTyrAspPheTyrProGlnPheSerIleProValArgThr 105  
QY 352 GAATCTGAA---GCTGCATATCAACAATTTAATCCAAAAACACCGATAACAATGATACT 408  
DB 106 GluLeuGluPheTyrAlaArgGlyLysAlaAspSerLysTyrAsnValAspLysAspSer 125  
QY 409 GATTAATGTTGATATATAACATTTTGCATTTATCTCGTAAAGATGCAATCGAAGATCAG 468  
DB 126 TrpSerGlyGlyTyrTrpArg-----AspAsp--- 134  
QY 469 CAATATGATGACTTAAAAATGACGGCATAACTTTTATGTCTATGATGGTTTAAATCTTGC 528  
DB 135 -----LeuLysAsnGlu---ValSerValAsnThrLeuMetLeuAsnAlaTyr 149  
QY 529 TATGACATTAACAGTGAAGAGTATCTTGTACCATATGTCATGTCAGGATGATAGGACGA 588  
DB 150 TyrAspPheArgAsnAspSer---AlaPheThrProTrpValSerAlaGlyIleGly--- 167  
QY 589 GATCTTATCACTATTTTAAAGACTCAATCTAAAA----- 624  
DB 168 -----TyrAlaLysGluIleHisGlnLysThrThrGlyIleSerThrTrpAsp 183  
QY 625 -----TTT 627  
DB 184 TyrGlyTyrGlySerSerGlyArgGluSerLeuSerArgSerGlySerAlaAspAsnPhe 203  
QY 628 GCTTACCAGAAAATAGGTATTAGTTTACCTATCACACCAGAGTCTCT 678  
DB 204 AlaTrpSerLeuGlyAlaGlyValArgTyrAspValThrProAspIleAla 220

## RESULT 10

AH0541  
probable outer membrane adhesin STY0351 [imported] - Salmonella enterica subsp. enterica  
C;Species: Salmonella enterica subsp. enterica serovar Typhi  
A;Note: this species has also been called Salmonella typhi  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001  
C;Accession: AH0541  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A;Authors: Park, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi  
A;Reference number: AB0502; PMID:11677608  
A;Accession: AH0541  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-239 <PAR>



```
QY 589 GATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGGCTTACCAAGGAAATAGGT 648
Db 421 ThrAspPheIleIleAspSerValAsnIleArgProIleTyrAsnGlyAlaTyrLys 440
QY 649 ATTAGTTACCTTATCACACCAAGTCTCTGCATTTATTGGTGGATACATACCATGGCGTT 708
Db 441 HisTyrTyrValValGlyAlaHisGlnSer-----TyrHisGly--- 453
QY 709 ATTGGTAATAATTGAGAAGATACCTGTATAACTCTGTAGTATTAAATGATGCTCCT 768
Db 454 -----PheGlu-----AspThrPro 458
QY 769 CAA-----ACCACATCTGCTTCACTAACTCTTGAC 798
Db 459 ArgArgArgIleThrLysSerAlaSerPheThrValAsp 471

RESULT 12
D90129
hypotheical protein orfi1613 [imported] - Guillardia theta nucleomorph
C:Species: nucleomorph Guillardia theta
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: D90129
R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Reil
Nature 410, 1091-1096, 2001
A>Title: The highly reduced genome of an enslaved algal nucleus.
A:Reference number: A99082; MUID:11323671; PMID:11323671
A:Accession: D90129
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1613 <DOU>
A:Cross-references: GB:AF083031; NID:gl3794350; PIDN:AAK39727.1; GSPDB:GN00152
C:Genetics:
A:Gene: Orfi1613
A:Map position: 3
A:Genome: nucleomorph
C:Keywords: nucleomorph

Alignment Scores:
Pred. No.: 0.306 Length: 1613
Score: 103.00 Matches: 58
Percent Similarity: 38.93% Conservative: 51
Best Local Similarity: 20.71% Mismatches: 75
Query Match: 7.16% Indels: 96
DB: 2 Gaps: 14

US-10-062-624-41 (1-840) x D90129 (1-1613)

QY 753 TACTACAGAGTATTACAGG-----TATCTT 727
Db 1169 TyrGluLysAsnTyrTyrLysThrLysSerThrCysLysTyrPhePheLysTyrPheVal 1189
QY 726 CTCAAA-----TTTATTACCAATACGCGCATGGTAGTATCCACCAATAAATGCAGA 676
Db 1189 LeuLysArgIlePheAspIleAsnGlnIleLeuPheIleAlaThrAsnLys----- 1206
QY 675 GACTTCTGGTGTAGAGGTAACTAACTATTTTCTGGTAAAGCAATTTAGATT 616
Db 1207 -----LeuLeuPheAspIleValAsnArgPhe---Ile 1216
QY 615 GAGGTCTTTAAAATAGTATAGATCTGCTCTCTACCTGACATGCATA----- 565
Db 1217 GluIleLeuAsnAsnLeuAsnLysIleIleThrTyr-----CysValIleLysLeu 1233
QY 564 -----TGGTACGAA 556
Db 1234 AsnLysLysThrGluLysAsnTyrTyrArgPheAsnPheValTyrHisSerPhePheGlu 1253
QY 555 AGA-----TACTCTTCAGCTGT 538
Db 1254 ArgAspLeuTyrAsnPhePheLeuTyrLeuSerIleGlnMetLysSerPhePheCys 1273
QY 537 AATGTCATAGCAAGTATTAAACCATCAATGACATAAAGTTATGCGCGTCATTTTAAAG--- 481
```

```
Db 1274 SerHisLeuSerSerLysAsnIleAsnAsnSerThrValThrLysValCysLysLysLys 1293
QY 480 -----TACTACATATTCGTATCTTCATTCCTGATCTTTACAGATAATGCAAA 433
Db 1294 LysSerLysSerHisPheIleArgIleLysLeuAsnLysIleTyrAspArgPheAspLys 1313
QY 432 ATGTTTATAGTATTTCACCATATCAGTATCATCTGTTATCGGTGTT-----TTTTGGATT 379
Db 1314 -----ThrIleLeuHisThrIleTyrSerValThrAsnTyrPheIle 1328
QY 378 AAATTGTTGATATGACG-----TTCAAGTTTCTATTCTGTCGTCCTCATAGAGTA 328
Db 1329 LysLys-----CysAspLysTyrMetLeuMetPheTyrLysPheLeuMetLysArgIle 1346
QY 327 ACCAATATCTCTCTGAAATCTCTGATATTAAGTTATTTGAAAATCAATGCTCGGAGCTAC 268
Db 1347 Phe-----LeuAspSerValAsnTyr 1353
QY 267 TCTTGTAAATC-----GTCTTTTGTATTATATCACCATCTTTCTTTAGTCCGAA 217
Db 1354 AsnTyrLysIleSerTyrGluIlePhePheGlnIleSerValIleSerPheSer 1373
QY 216 AACTTTTCTAGTGAGAGAAATTTGTTCCATTATAGAGTTTCTTCAGCAGAGAAATTTCT 157
Db 1374 AsnLeuPheProLysThrAsnLeuLysIleMetLysAsnLeuPheIleLysLysPheAsn 1393
QY 156 AAAGTGTATATATCTGGATTGTACTTTGGCACTAATGTAGAGGCTTCTTTGTTATCAAT 97
Db 1394 LeuIleAsnLeuIleIleLeuLysGluLeuArgAsnPheSerAlaThrPheLeuIleLeu 1413

RESULT 13
A27826
DNA-directed RNA polymerase (EC 2.7.7.6) II second largest chain - fruit fly (Drosophila
C:Species: Drosophila melanogaster
C>Date: 19-May-1989 #sequence_revision 23-Nov-1991 #text_change 21-Jul-2000
C:Accession: A27826; PQ0154
R:Falkenburg, D.; Dworniczak, B.; Faust, D.M.; Bautz, E.K.F.
J. Mol. Biol. 195, 929-937, 1987
A>Title: RNA polymerase II of Drosophila. Relation of its 140,000 Mr subunit to the beta
A:Reference number: A27826; MUID:88011299; PMID:3116266
A:Accession: A27826
A:Molecule type: DNA
A:Residues: 54-1176 <PAL>
A:Cross-references: GB:X05709; GB:M29646; NID:G5514651; PIDN:CAA29180.2; PID:G5514652
R:Sitzler, S.; Oldenburg, I.; Petersen, G.; Bautz, E.K.F.
Gene 100, 155-162, 1991
A>Title: Analysis of the promoter region of the housekeeping gene DmrP140 by sequence co
A:Reference number: PQ0154; MUID:91276237; PMID:1905256
A:Accession: PQ0154
A:Molecule type: mRNA
A:Residues: 1-69 <SIT>
A:Cross-references: GB:MG2972; NID:gl57263; PIDN:AAA28476.1; PID:gl57267
A:Experimental source: embryo
C:Genetics:
A:Gene: DmrP140
A:Cross-references: FlyBase:FBgn0003276
A:Introns: 8/1
C:Superfamily: DNA-directed RNA polymerase 132K polypeptide
C:Keywords: DNA binding; nucleotidyltransferase; nucleus; transcription; zinc finger

Alignment Scores:
Pred. No.: 0.365 Length: 1176
Score: 102.00 Matches: 53
Percent Similarity: 34.59% Conservative: 39
Best Local Similarity: 19.92% Mismatches: 78
Query Match: 6.89% Indels: 96
DB: 2 Gaps: 12

US-10-062-624-41 (1-840) x A27826 (1-1176)

QY 61 TATCAGTCTTTTGCAGATCTCTGTAGTTCAAGAACTATGATAACAAAGAGGCTTCTTAC 120
```



RESULT 15  
AC1153  
adhesin homolog lmo0627 [imported] - Listeria monocytogenes (strain EGD-e)  
C:Species: Listeria monocytogenes  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AC1153  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Jones, L.M.; Karst, U.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Science 294, 849-852, 2001  
A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Makok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AC1153  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1582 <GLA>  
A:Cross-references: GB:NC 003210; PID:NC98705.1; PID:gl6410016; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo0627

Alignment Scores:  
Pred. No.: 0.572 Length: 1582  
Score: 100.00 Matches: 63  
Percent Similarity: 31.53% Conservative: 30  
Best Local Similarity: 21.36% Mismatches: 84  
Query Match: 6.75% Indels: 118  
DB: 2 Gaps: 13

US-10-062-624-41 (1-840) x AC1153 (1-1582)

QY 115 TTCTACATTATGCGAAGTACCAAGTATATACACTTTAGAAAATCTCTCTGAA 174  
DB 548 PheTyrMetAlaGlyGlu-----IleGlySerPheSerAsnPheAsnGlyGlu 563  
QY 175 GAACTCTTATTAATGGACAAATCTCTCACTAAAGATTTTCGGACTAAGAAAGAT 234  
DB 564 SerThr-----LeuAsp 567  
QY 235 GGTGATATACAAAAAGACGATTTTACAGAGTAGCTCCAGGCATTGATTTTCAAAAT 294  
DB 568 GlyAlaIleThrAspValAspSerThrProAspMetAsnProGly-----Asn 583  
QY 295 AACTTAATATACAGATTTTCAGGAAGTAT-----GGTACTCTATGAC 339  
DB 584 AspLeuIleSerAsnValAspGlySerThrIleLysProIleAspGlyPheGlnLysGlu 603  
QY 340 GGACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACCCGATAAC 399  
DB 604 LysProHisValValLysGluAsnAlaLysSerThrAlaThrLeuGlnAspGluAspAsp 623  
QY 400 AATGATACCTGAT-----AATGGTGA 420  
DB 624 PheAspPheAspTyrValThrPheIleLysThrProGluValIleProSerGluGlyLys 643  
QY 421 TACTATAACATTTTGCATTTCTCGTAAAGATGCAATGGAAGAT----- 465  
DB 644 ValPheGluLysThrArgLysSerAlaAlaAspAlaGluLysThrAlaArgAsnMetIle 663  
QY 466 -----CAGCAATATGATAGTACTTTAAAAATGACCGC----- 495  
DB 664 GlyAsnAspIleMetGlySerAlaLeuLysAspTyrThrProLeuAlaAspAspGlyArg 683  
QY 496 -----ATACTTTTATGCTATGATGGTTAAT----- 522  
DB 684 IleThrSerProLysThrTyrMetValTyrGluMetAsnIleAsnProSerGlyValGlu 703  
QY 523 -----ACTTGCTATGACATTACAGCTGAAGAGTATCTTTCTGACCATAT 567  
DB 704 AspThrLeuAsnSerSerPheThrAspThrLeuProLysGlyLeuLysMetLeuGluTyr 723

QY 568 GCATGTGCAGGTATAGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTT 627  
DB 724 AspIleAlaGlyLeuAsnSerSer----- 731  
QY 628 GCTTACCAAGAAATAAGTATTAGTTACCTATACACCAAGAGTCTCTGCATTTATT 687  
DB 731 ----- 731  
QY 688 GGTGATACTACCATGGCTTTATTTGGTAAATAAATTTGAGAAGATACCTGTAAATACT--- 744  
DB 732 AsnLysHisIleTyrGlyPheThrThrAsnLysPheGluGlyThrProLysPheThrAsn 751  
QY 745 -----CCTGTAGTA-----TTAAATGATGCTCTCAACACCACA 777  
DB 752 AlaAspGlyGlnAspValIleValTyrGlnLysGlyLeuTyrSerGluLysGlnValCys 771  
QY 778 TCTGCTTCAGTAACTCTT---GAGCTTTGGA-----TACTTTGGC 813  
DB 772 IleAlaAsnProThrLeuAsnAspValGlyValArgTyrPheGly 786

#### RESULT 16

C89921  
hypothetical protein ebhB [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: C89921  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc, ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K. Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:213111952; PMID:11418146  
A:Accession: C89921  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-3890 <KUR>  
A:Cross-references: GB:BA000018; PID:gl3701233; PID:BA042528.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: ebhB

Alignment Scores:  
Pred. No.: 0.953 Length: 3890  
Score: 98.00 Matches: 65  
Percent Similarity: 38.83% Conservative: 48  
Best Local Similarity: 22.34% Mismatches: 114  
Query Match: 6.62% Indels: 64  
DB: 2 Gaps: 14

US-10-062-624-41 (1-840) x C89921 (1-3890)

QY 7 TATAAGAAATTTAGTAAAGCGGTAAATCTCAATTAATGTCAATCTTACCATATCAG 66  
DB 892 TyrAsnLysVal---ValAlaSerAsnMetSerSerAlaValThrIleLeuProAspAsp 910  
QY 67 -----TCTTTTGCAGATCCTGTAGTTCAACAACATATGATAACAAGAGGCTTC 117  
DB 911 IleProThrIleAsnAsnProValGly----- 920  
QY 118 TACATTAGTCGAAAGTACAATCCAAAGTATATACACTTTAGAAAATTTCTGCTGAAGAA 177  
DB 921 ---IleAsnAlaLysTyr-----TyrArgGlyAspGlu 930  
QY 178 ACTCTTATTAAGCAACAATCTCTCACTAAAAAGTTTTCGGACTAAGAAAGATGCT 237  
DB 931 -----ValAsnPheThrMetGlyValSerAspArgHisSerGlyIleLysAsnThrThr 948  
QY 238 GATATAACAAAAAAGACGATTTTACAGAGTAGCTCCAGGCATTGATTTTCAAAATAAC 297  
DB 949 IleThrThrLeuProSerGlyTyrPheThrSerAsnLeuThrLysSerAspAsnLysAsn--- 967  
QY 298 TTAATATCAGGATTTTTCAGGAAGTATTGGTTACTCTATGACGGACCAAGATAGAACTT 357  
DB ----- 357

```

Db 968 -----GlySerLeuAlaIleThr-----GlyArgValSerMet 978
Qy 358 GAAGCTGATATCAACAA-----TTTAATCCAAAAAACACCGATAAC---AATGAT 405
Db 979 AenGlnAlaPheAenSerAspIleThrPheLysValSerAlaThrAspAenValAenAen 998
Qy 406 ACTGATATGCTGAATACTATAACATTTTCATTAATCTCTGTAAGATGCAATGGAAGAT 465
Db 999 ThrThrAenAenSerGlnSerLysHisValSerIleHisValGlyLysIleSerGluAasp 1018
Qy 466 CAGCAATATGATGATCTTAAAAATGACGGCATACTTTTATGTCATTGATGTTAAATACT 525
Db 1019 AlaHisProIleValLeuGlyAen-----ThrGluLysValValValValAenPro 1035
Qy 526 TGCTATGACATTACAGCTGAAGAGTATCTTTCGTACCATATGTCATGTCAGGATATAGGA 585
Db 1036 ThrAlaValSerAenAenGluLysGlnSer-----1045
Qy 586 GCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAATA 645
Db 1046 -----IleIleThrAlaPheMetAenLysAenGlnAenIle-----ArgGlyTyrLeu 1061
Qy 646 GGTATTAGTACCTTATCACACAGAGTCTCTGCATTATTGCTGGATACCTACCATGGC 705
Db 1062 AlaSerThrAaspProValThrValAaspAenAenGlnAenValThrLeuHisTyrArgAasp 1081
Qy 706 GTTATTGGTAATAAATTGGAAGATACCTGTAATAACT-----CCTGTAGTATTAAAT 759
Db 1082 GlySerSerThrThrLeuAenAlaThrAenValMetThrTyrGluProValValLysSer 1101
Qy 760 GAT-----GCTCTCAACACCATCTGCTTCAGTAACCTCTTGACGTTGGA 804
Db 1102 GluTyrGlnThrAlaAenAlaAlaLysThrAlaThrValThrIleAlaLysGlyGlnSer 1121
Qy 805 TACTTTGCGGAGAAATTTGGAATCAGGTTCACC 837
Db 1122 PheAenIleGlyAspIleLysGlnTyrPheThr 1132

RESULT 17
AC3295
heat resistant agglutinin 1 precursor [imported] - Brucella melitensis (strain 16M)
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C;Accession: AC3295
R;DeiVecchio, V.G.; Kapral, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Leteser
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A;Reference number: AD3252; PMID:11756688
A;Accession: AC3295
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-274 <KUR>
A;Cross-references: GB:AE008917; PIDN:AAU51526.1; PID:gl7982243; GSPDB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI0345
A;Map position: 1

Alignment Scores:
Pred. No.: 0.807 Length: 274
Score: 97.50 Matches: 69
Percent Similarity: 31.60% Conservative: 34
Best Local Similarity: 21.17% Mismatches: 98
Query Match: 6.58% Indels: 125
DB: 2 Gaps: 17

US-10-062-624-41 (1-840) x AC3295 (1-274)
Qy 7 TATAAGAAATTTCTAGTAAGAGCGCTTAATCTCAATTAATGTCATC-----54
Db 8 TyrAlaGlyLeuIleAlaSerAlaAlaLeuIleValSerPheSerValAlaSerAlaThr 27

```

```

Qy 55 -----TTACCATATCACTCTTTTGCAGATCCT---GTAGGTTCAAGAACT 96
Db 28 AspLeuProGluThrValProGluValMetIleAlaProProAlaValGly-----44
Qy 97 AATGATAACAAGAAGGCTTCTACATTTAGTGCAAAGTACAACTCAAGTATATCACACTTT 156
Db 45 -----GlyTyrTyrLeuArgGlyAaspIleGlyTyrSerTyrPheGlyPhe 59
Qy 157 AGAAAATTTCTGCTGAAGAACTCTTATTAATAGGAACAATACTCTCTCACTAAAAAAGTT 216
Db 60 LysArg-----AlaGluTyrAlaProLeuGlyAaspCysAaspThrCysAen-----74
Qy 217 TTGCGACTAAAGAAGATGCTGATATAACAAAAAAGAC-----255
Db 75 ---GlyLeuGlyArgGlyGlyAaspThrLeuTyrGlyAaspLeuAenGlySerPheLeuIle 93
Qy 256 -----GATTTTACAAGAGTAGCTCCAGGCAATTTT 288
Db 94 GlyGlyGlyAlaGlyTyrGlnValThrAaspTyrPheArgThrAaspLeuThrLeuAaspTyr 113
Qy 289 CAAAATAACTTAATATCAGGATTTTTCAGGAAGTATTGTTACTCTATGCGACGGACCAAGA 348
Db 114 MetThrArg-----SerArgPheSerGlyHisVal-----SerGlyProAasp 127
Qy 349 ATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACCCGATAACAATGATACT 408
Db 128 Cys-----GNTAATGCTGAATACTATATAACAATTTTGCATTTATCTCTGTAAGATGCA 456
Qy 409 -----GATTAATGCTGAATACTATATAACAATTTTGCATTTATCTCTGTAAGATGCA 456
Db 133 GlyCysAlaSerAaspGluArgSerHisTyr-----142
Qy 457 ATGGAAGATCAGCAATATAGTAGTACTTAAAAATGACGGCATAACTTTTATGTCATTGATG 516
Db 143 -----SerAlaLeuSerIleLeu 148
Qy 517 GTTAATACTTGTATGATACATTACAGCTGAAGAGATATCTTTTCGTACCATATGTCATGCA 576
Db 149 AlaAenAlaTyrValAaspLeuGlyAenLeuGly---GlyValThrProTyrValGlyAla 167
Qy 577 GGTATAGGA-----GCAGATCTTATCACTATTTTAAAGACCTC---615
Db 168 GlyIleGlyGlyThrArgValAenTrpSerAaspLeuValAaspIleThrSerGlyPheSer 187
Qy 616 -----AATCTAAATTTGCTTACCAAGAAATAATAGTATTAGTTACCTCCT 660
Db 188 GlnGluGlyAlaAlaAenTrpArgPheThrTyrAlaLeuMetAlaGlyAlaSerValAasp 207
Qy 661 ATCACACCAGAGTCTCTGCAATTTATGTTGGATACCTACCATGCGGTATTGTTGTAATAA 720
Db 208 LeuThrHisAenLeuLysLeuAaspAlaGlyTyrArgTyrArgHisValAenGlyGlyLys 227
Qy 721 TTTGAGAGATACCTGTAATACTCTGTAGTAGTATTAATGATGCTCTCTCAACCATCATCT 780
Db 228 -----MetPheGluGlyAenGlnTrpThr---235
Qy 781 GCTTCAGTAATCTCTGACGTTGATACCTTTGGCGGA-----GAAATT 822
Db 236 -----AspAlaGlyTyrAaspLysGlyLeuAenIleHisAaspIleArgVal 250
Qy 823 GGAATGAGGTTTCACCTTC 840
Db 251 GlyLeuArgTyrMetPhe 256

RESULT 18
S28976
DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain - human
N;Alternate names: DNA-directed RNA polymerase B largest chain; DNA-directed RNA polymerase
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 18-Jun-1999
C;Accession: S28976; S18986
R;Ackner, J.; Wintzerith, M.; Vigneron, M.; Keding, C.
J. Mol. Biol. 226, 1295-1299, 1992

```

A:Title: Primary structure of the second largest subunit of human RNA polymerase II (or

A:Reference number: S28976; MUID:92389336; PMID:1518060

A:Accession: S28976

A:Molecule type: mRNA

A:Residues: 1-1174 <ACK>

A:Cross-references: EMBL:X63563; NID:g36121; PIDN:CAA45124.1; PID:g36122

C:Superfamily: DNA-directed RNA polymerase 132K polypeptide

C:Keywords: nucleotidyltransferase; transcription

Alignment Scores:  
Pred. No.: 0.936 Length: 1174  
Score: 97.50 Matches: 57  
Percent Similarity: 32.96% Conservative: 31  
Best Local Similarity: 21.35% Mismatches: 92  
Query Match: 6.58% Indels: 87  
DB: 2 Gaps: 11

US-10-062-624-41 (1-840) x S28976 (1-1174)

```
Qy 115 TTCTACATTAGTGCAGTACCAATCCAGTATATACACACTTATAGAAAATTCCTGCTGAA 174
Db 810 PheTyrArgSerTyrLysGluGlnGluSerLysLysGlyPheAspGlnGluValPhe 829
Qy 175 GAAACTCCTATTAAATGGAACAAATCTCTCACTAAAAAGTTTCGGACTAAAGAAAGAT 234
Db 830 GluLysPro-----ThrArgGluThrCysGlnGlyMetArgHis 842
Qy 235 GGTGATATAACAAAAAGACGATTTTACAAGAGTAGCTCCAGGCATTGATTTTCAAAAT 294
Db 843 AlaIleTyrAspLysLeuAspAspGlyLeuIleAlaProGlyVal----- 858
Qy 295 AACTTAATATCAGGATTTTCAGGAGTATTGGTTACTCTATGCGCGACCAAGATAGAA 354
Db 859 ---ArgValSerGlyAspValIleIleGlyLysThrValThr----- 872
Qy 355 CTGGAAGCTGCATATCAACAATTTAATCCAAAAACCGGATAACAATGATGATAAT 414
Db 873 -----LeuProGluAsnGluAspGluLeuGluSerThrAsn 884
Qy 415 GGTGAATATATAACAT-----TTTGCAATTTATCTCGTAAAGATGCAATGAA 462
Db 885 ArgArgTyrThrLysArgAspCysSerThrPheLeuArgThrSerGluThrGlyIleVal 904
Qy 463 GATCAGCAATATGATAGTACTTAAATACACGCATAACTTTTATGCTATTGATGGT--- 519
Db 905 AspGlnValMetValThrLeuAsnGlnGluGlyTyrLysPheCysLysIleArgValArg 924
Qy 520 -----AATACT 525
Db 925 SerValArgIleProGlnIleGlyAspLysPheAlaSerArgHisGlyGlnLysGlyThr 944
Qy 526 TGCTATGACATTACAGCTGAAGAGTATCTTTC-----GTACCATATGCAATGT 573
Db 945 Cys-----GlyIleGlnTyrArgGlnGlnGluAspMetProPheThrCys 958
Qy 574 GCAGGTATAGGACGACATCTTATCACT----- 600
Db 959 GluGlyIleThrProAspIleIleIleAsnProHisAlaIleProSerArgMetThrIle 978
Qy 601 -----ATTTTAAAGACTCAATCTAAATTT---GCTTACCAAGGAAAAATAGGTATT 651
Db 979 GlyHisLeuIleGluCysLeuGlnGlyLysValSerAlaAsnLysGlyGluIleGlyAsp 998
Qy 652 AGTTACCTATC-----ACACCAAGAGTCTCTGCAATTTATTGGTGGATAC 696
Db 999 AlaThrProPheAsnAspAlaValAsnValGlnLysIleSerAsnLeuLeuSerAspTyr 1018
Qy 697 -----TACCATGGCGTTATTGGTAAATAAATTGAG 726
Db 1019 GlyTyrHisLeuArgGlyAsnGluValLeuTyrAsnGlyPheThrGlyArgLysIleThr 1038
Qy 727 AAGATACCTGTAAATCACTCT 747
```

Db 1039 SerGlnIlePheIleGlyPro 1045

RESULT 19

A45555

glutamate rich protein - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 09-Jun-2000

C:Accession: A45555; S27831

R;Borre, M.B.; Driegiel, M.; Hogh, B.; Petersen, E.; Rieneck, K.; Riley, E.; Meis, J.F.; J.

Mol. Biochem. Parasitol. 49, 119-131, 1991

A:Title: Primary structure and localization of a conserved immunogenic Plasmodium falciparum

rate life cycle.

A:Reference number: A45555; MUID:92131041; PMID:1775153

A:Accession: A45555

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1271 <BOR>

A:Cross-references: EMBL:M59706; NID:g160311; PID:g160312

A:Note: sequence extracted from NCBI backbone (NCBIN:77801, NCBI:77802)

Alignment Scores:  
Pred. No.: 0.944 Length: 1271  
Score: 97.50 Matches: 54  
Percent Similarity: 40.64% Conservative: 35  
Best Local Similarity: 24.66% Mismatches: 83  
Query Match: 6.58% Indels: 47  
DB: 2 Gaps: 11

US-10-062-624-41 (1-840) x A45555 (1-1271)

```
Qy 22 GTAAGACCGGTAAATCTCATTAATCTTACCATATCAGTCTTTTCAGATCCT 81
Db 1079 ValGlnHisGluIleValGluValGluLeuProGluIleValGluIleGluGlu 1098
Qy 82 GTAGGTTCAAGAACTAATGATAACAAA-----GAAGGCTTCTACATTAGTGCAGGATAC 135
Db 1099 ValProSerGlnThrAsnAsnGluAsnIleGluThrIleLysProGluGluLysLys 1118
Qy 136 AATCAAGATATATACACTTTAGAAAATTTCTGCTGCAAGAA-----ACT 180
Db 1119 Asn-----GluPheSerValGluGluLysAlaIleProGlnGlu 1131
Qy 181 CCTATTATGGAACAAATTTCTCCTACTAAAAAGTTTCGACTAAAGAAAGATGAT 240
Db 1132 ProValValProThrLeuAsnGluAsnGluAsnValThrProLysProSerGluGluGlu 1151
Qy 241 ATAACAAAAAAGACGATTTTACAAGAGTAGCTCCAGGCATTGATTTTCAAAAATAACTTA 300
Db 1152 SerThrLysProAsp-----IleValGlnIleLysIleValGlnGluAsnLys 1167
Qy 301 ATATCAGGATTTTCAGGAAGTATTGGTTACTCTTATGACGCGACCAAGAAAGATGAA 360
Db 1168 ProAsnLysLysGluThrProVal-----ValAspGlyProLys---HisValGlu 1183
Qy 361 GCTGCATATCAACAATTTAATCCAAAAACACCGATCAATGATGATTAATGTTGAA 420
Db 1184 GlnAsnIleGlnGlu-----AspAspAsnAspGluGluAspAspAsp 1197
Qy 421 TACTATAAACAATTTTGCATTATCTCGTAAAGATGCAATGCAAGATCAGCAA----- 471
Db 1198 AspileAspPheGluGlyLeuSerArgLysAspAspGluLysAspSerSerAsnLysAsn 1217
Qy 472 -----TATGTA-----GTACTTAAATAATGACGGCATA 498
Db 1218 LysLysLysSerSerPheIleThrTyrIleSerThrLysLysPheLysLysValSerGln 1237
Qy 499 ACTTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 558
Db 1238 ThrIleValSerValMetIleAsn---AlaTyrAsp-----GlyValIleGln 1252
Qy 559 GTACCATATGATGTCGAGGTATAGGACGATCTTATCACTATTTTAAAGACCTC 615
```



Db 1253 ValValSerThrIleLysGlyLeuAlaLysAspIleValIlePheGluAsnIle 1271

RESULT 20  
AD1380  
glycosidase homolog lmo2444 [imported] - Listeria monocytogenes (strain EGD-e)  
C:Species: Listeria monocytogenes  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AD1380  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Mok, C.; Schluecker, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AD1380  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1310 <GLA>  
A:Cross-references: GB:NC 003210; PIDN:CAD00522.1; PID:g16411932; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo2444

Alignment Scores:  
Pred. No.: 0.947 Length: 1310  
Score: 97.50 Matches: 76  
Percent Similarity: 31.08% Conservative: 25  
Best Local Similarity: 23.38% Mismatches: 131  
Query Match: 6.58% Indels: 93  
DB: 2 Gaps: 15

US-10-062-624-41 (1-840) x AD1380 (1-1310)

Qy 88 TCAGAGACTATGATAAC-----AAAGAGGCTTCTACATTAGT 126  
Db 791 SerLysThrAlaAspThrSerThrPheThrIleGlyAlaLysGluGlySerTyr---Lys 809  
Qy 127 GCAGAGTACATCAAGTATATACACCTTTAGAAAATCTCTGCTGAAGAACTCCCTATT 186  
Db 810 AlaAspThrGluAsnTyrIleValLysLeuHisGlyAsnAlaAlaGluSerValGlnVal 829  
Qy 187 AATGAACAAATCTCTCACTAAAAGATTTC----- 219  
Db 830 AsnGlyAlaAspSer-----LysValTyrAspAspLeuSerAlaLeuGluAsnAla 846  
Qy 220 -----GGACTAAAGAGATGGTGATATACAAAAGAGAGATTTTACAGAGTA 270  
Db 847 AspGlySerGlyTyrAlaValGlySerAspValTyrGlyAlaValThrTyrValLysVal 866  
Qy 271 GCTCAGGCAATGATTTTCAAAATAACTTAATATCAGGATTTTCAGGAAGTATT----- 324  
Db 867 ProAlaGlnAlaAspAlaAspThrValIleThrThrSerGlySerAlaSerIleValGln 886  
Qy 324 ----- 324  
Db 887 ThrGlyThrTyrGluMetGluThrGlySerAsnPheAlaAspThrValAlaAspLysPro 906  
Qy 325 -----GGTTACTCTATGACCGACCAAGATAGAACTTGAA 360  
Db 907 ValAlaGluLysThrTyrValAspGlyTyrAspLysAspGlyAlaGlyThrThrIleTyr 926  
Qy 361 GCTCATATCAACAATTAATCCAAAACACCGCAT-----AACAAATGACT 408  
Db 927 AlaAsnValLysAspSerGlyAspTyrAsnValAspLeuThrTyrLysAsnAlaSerSer 946  
Qy 409 GAT-----AATGGTGAATATATAACATTTTGCATTATCT 444  
Db 947 AspAsnGlnAlaLeuSerIlePheValAsnGlyGluTyrValLysGlnThrThrLeuLys 966  
Qy 445 CGTAAAGATGCAATGGAGATCAGCAATATGATCTTAAATATGACCGCATACCTTTT 504  
Db 445 CGTAAAGATGCAATGGAGATCAGCAATATGATCTTAAATATGACCGCATACCTTTT 504

Db 967 ProAsnThrAspTyrPsrValGlnSerGluThrLeuProLeuSerAlaGlyLysAsnSer 986  
Qy 505 ATGTCAATTGATGGTTAATACTTGTATGACATTACAGCTGAAGAGATATCT----- 555  
Db 987 IleSerTyrLysValValThrAspThrGlyAspLysAlaAspGlnValSerLeuAspLys 1006  
Qy 556 -----TTGCTACCATATGCAATG-----GCAGGTATAGGAGAGATCTTATC 597  
Db 1007 ValAsnIleGlyPheThrProThrValAlaLysValGluAlaGluGluAlaAspLeuAla 1026  
Qy 598 ACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAGGAAATAAGTATTAGTTAC 657  
Db 1027 GlyThrLeuLysValAlaAsnAspHisThrPheTyrSerGlyGlu----- 1041  
Qy 658 CCTATCACACAGAGTCTCTGCATTTATTTGGTGATATACCATCGCGTATTGGTAAT 717  
Db 1042 -----GlyPheValGlyGlyPhe-----GlyThrAlaGlyAsp 1052  
Qy 718 -----AAATTTGAG-----AAGATACCTGTAACTCTCTGTATTTAAATGAT---GCT 765  
Db 1053 GluIleLysPheGluValAspValProAlaAspGlyGluTyrValLeuAsnThrArgThr 1072  
Qy 766 CCTCAACACCATCTGCTTCACTAACTCTTGAGCTTGAGTACTTTGCGGAGAAATTGGA 825  
Db 1073 AlaAsnGlyThrGlyMetProGlnThrLeuAspLeu---TyrThrAsnGlyAspTyrAsn 1091  
Qy 826 ATGAGGTTCACTTC 840  
Db 1092 SerArgValThrPhe 1096

RESULT 21  
D90092  
hypothetical protein orf665 [imported] - Guillardia theta nucleomorph  
C:Species: nucleomorph Guillardia theta  
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001  
C:Accession: D90092  
R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Rel.  
Nature 410, 1091-1096, 2001  
A:Title: The highly reduced genome of an enslaved algal nucleus.  
A:Reference number: A99082; MUID:11323671; PMID:11323671  
A:Accession: D90092  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-665 <DOU>  
A:Cross-references: GB:AF165818; NID:g13794493; PIDN:AAK39868.1; GSPDB:GN00150  
C:Genetics:  
A:Gene: orf665  
A:Map position: 1  
A:Genome: nucleomorph  
C:Keywords: nucleomorph

Alignment Scores:  
Pred. No.: 1.49 Length: 665  
Score: 95.00 Matches: 66  
Percent Similarity: 39.31% Conservative: 59  
Best Local Similarity: 20.75% Mismatches: 91  
Query Match: 6.60% Indels: 102  
DB: 2 Gaps: 15

US-10-062-624-41 (1-840) x D90092 (1-665)

Qy 768 AGGAGCATATTAATACTACAGAGTTAT---TACAGGTATCTTCTCAATTTATTATACC 712  
Db 362 ArgAsnIleIleIleTyrLeuArgThrTyrSerPheValTyrCysIleLysPheLeuLeu 381  
Qy 711 AATAACGCCATGGTAGTATCCCAATAATGCAGAGACTCTTGGTGTGTAGAGTAACT 652  
Db 382 IleAsnProArgIleIleLeuPheSerIleThrLysGlu-----IleLys 396  
Qy 651 AATACCTATTTTCTTGGTGAACAAATTTTAGATTGAGGTCTTTAAATAATAGTGATAAG 592  
Db 397 AsnIlePheGluIleIleSer-----AsnAsnAsnIle 408



Db 768 GluGlyTyrThrIleAlaGlyValGluThrLysGlyGluValValThrIleLysValThr 787  
QY 745 CTGTAGTATTAAATGATCTCTCTCAACACCATCTGCTTCACTA 789  
Db 788 GluLeuAspIleAspAspSerAspAlaThrProThrValAlaVal 802  
RESULT 23  
AD1512  
peptidoglycan bound protein (LPXTG motif) adhesin homolog lin0636 [imported] - Listeria  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AD1512  
R:Glaeser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A.:Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; PMID:21537279; PMID:11679669  
A:Accession: AD1512  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1578 <GLA>  
A:Cross-references: GB:ALU592022; PIDN:CAC95868.1; PID:g16413076; GSPDB:GN00178  
A:Experimental source: strain Clp11262  
C:Genetics:  
A:Gene: lin0636  
Alignment Scores:  
Pred. No.: 1.63 Length: 1578  
Score: 95.00 Matches: 64  
Percent Similarity: 31.77% Conservative: 31  
Best Local Similarity: 21.40% Mismatches: 86  
Query Match: 6.41% Indels: 118  
Gaps: 14  
DB:  
US-10-062-624-41 (1-840) x AD1512 (1-1578)  
QY 115 TTCTACATTAGTGAAGTACAATCAAGTATATCACACTTTAGAAAAATTTCTCTGCTGAA 174  
Db 548 PheTyrMetAlaGlyGlu-----lleGlySerPheSerAspPhe----- 560  
QY 175 GAAACTCTTATTAATGGAACAAATCTCTCACTAAATAAGTTTCGGACTAAAGAAAGAT 234  
Db 561 -----AsnGlyAspSerThrLeu-----Asp 567  
QY 235 GGTGATATAACAAAAAGACGATTTTACAGAGTAGTCTCCAGGCATTTGATTTTCAAAAT 294  
Db 568 GlyAlaIleThrAspValAspSerThrProAspMetAsnProGly-----Asn 583  
QY 295 AACTTAATATCAGGATTTTCAGGAAGTATT-----GGTTACTCTATGAC 339  
Db 584 AspLeuIleSerAsnValAspGlySerThrValLysProIleAspGlyPheGlnLysGlu 603  
QY 340 GGACCAAGATAGACTTGAAGTCCATATCAACAATTTAATCCAAAAACACCGATAAC 399  
Db 604 LysProHisValLysGluAsnAlaLysSerThrAlaThrLeuGlnAspGluAspAsp 623  
QY 400 AATGATCTGAT-----AATGCTGAA 420  
Db 624 PheAspPheAspTyrValThrPheIleLysThrProGluValIleProThrGluGluLys 643  
QY 421 TACTATAACATTTTGGCATTTCTCGTAAAGATGCAATGGAGAT----- 465  
Db 644 ValPheGluLysThrArgLeuSerAlaAlaaspAlaGluLysThrAlaArgAsnMetIle 663  
QY 466 -----CAGCAATATGTAGTACTTTAAAAATGACGCG--- 495  
Db 664 GlySerAspIleMetGlySerAlaLeuLysAspTyrThrProLeuAlaAspGlyArg 683  
QY 496 -----ATAACTTTTATGTCATTTGATGTTAAT----- 522

Db 684 IleThrSerProLysThrTyrMetValTyrGluMetAsnIleAsnProSerGlyValGlu 703  
QY 523 -----ACTTGTATGACATTACAGCTGAAGGAGTATCTTCTGACCATAT 567  
Db 704 AspThrLeuAsnSerSerPheThrAspThrLeuProLysGlyLeuLysMetLeuGluTyr 723  
QY 568 GCATGTGCAGGTATAGGACAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTT 627  
Db 724 AspIleAlaGlyLeuAsnSerSer----- 731  
QY 628 GCTTACCAAGGAAAAATAGGTATTAGTTACCTATCACACCAGAGTCTCTGCATTATT 687  
Db 731 ----- 731  
QY 688 GGTGATATACCATGGCTTATTGGTAAATAATTTGAGAAGATACCTGTAATAACT--- 744  
Db 732 AsnLysHisIleTyrGlyPheThrAsnLysPheGluGlyThrProLysPheThrAsn 751  
QY 745 -----CCTGTAGTA-----TTAAATGATGCTCTCCAAACACCA 777  
Db 752 AlaAspGlyGlnAspValIleValTyrGlnLysGlyLeuTyrSerGluLysGlnValCys 771  
QY 778 TCTGCTTCAGTAACCTCTT---GACGTTGGA-----TACTTTGGCGGAGAAATTCGA 825  
Db 772 IleAlaAsnSerThrLeuAsnAspValGlyValArgTyrPheGlyAsnAsnAlaGly 790  
RESULT 24  
F97264  
6-Phospho-Beta-D-Galactosidase [imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C:Accession: F97264  
R:Nolling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: F97264  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-474 <KUR>  
A:Cross-references: GB:AB001437; PIDN:AAK80905.1; PID:g15026017; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC2963  
C:Superfamily: Agrobacterium beta-glucosidase  
Alignment Scores:  
Pred. No.: 1.6 Length: 474  
Score: 94.50 Matches: 59  
Percent Similarity: 35.27% Conservative: 38  
Best Local Similarity: 21.45% Mismatches: 107  
Query Match: 6.38% Indels: 71  
Gaps: 12  
DB:  
US-10-062-624-41 (1-840) x F97264 (1-474)  
QY 85 GGTTCAGAACTAATGATATAACAAAGAGGCTTCTACATTAGTGCAGAGTACAATCCAAGT 144  
Db 41 GlyArgPheThrGlyAspThrAlaSerAspPheTyr-----HisLysTyrLysGluAsp 58  
QY 145 ATATCACACTTTGAAAAATTTCTCTGCTGAAGAAACTCTTATTAATGGAACAAAATTTCTCTC 204  
Db 59 LeuLysPheSerArgLysPhe-----GlyValAsnGlyIle 70  
QY 205 ACTAAAAAGTTTTCGGACTAAAGAGATGGTGATATATAACAAAAAGACGATTTTACA 264  
Db 71 ArgIleSerIle-----AlaTrpSer 77  
QY 265 AGAGTAGCTCCA-----GGCATTGATTTTCAAAATAC 297  
Db 78 ArgValIleProAspGlyLysGlyGluValAsnProLysGlyLeuLysPheTyrSerAsp 97





C:Genetics:  
A:Map position: 5  
A:Note: FLN13\_10

## Alignment Scores:

Pred. No.: 2.29 Length: 745  
Score: 93.00 Matches: 70  
Percent Similarity: 34.60% Conservative: 48  
Best Local Similarity: 20.53% Mismatches: 107  
Query Match: 6.28% Indels: 116  
DB: 2 Gaps: 16

US-10-062-624-41 (1-840) x T51370 (1-745)

```

QY 43 TTAATGTCATCTTACCATATCAGTCTTTTCGAGATCTCTGAGTTCAAGAACTAATGAT 102
D 77 LeuSerProLeuProThrAsnSerPheGlnAsnPheThrLeuAsnAsnGlyAsp 96
QY 103 AACAAAGAGGCTCTAC-----ATTAGTCAAG 132
D 97 GlnAlaGluTyrPheHisProTyrIleLeuSerProSerThrSerLeuSerIleSer 116
QY 133 TACATCAAGTATATCACAC-----TTTAAAGAAATCTCTGCTGAAGAA 177
D 117 Tyr---ProSerLeuSerHisAsnSerAlaPheIleTyrGluAlaPheAsnAlaAspIle 135
QY 178 ACTGCTATTAATGNAACAAATCTCTCAATAAAAGTTTTCGGACTAAAGAAAGATGGT 237
D 136 Thr---IleThrGlySer-----AspGly 142
QY 238 GATATAACAAAAAGACGATTTTACAAGAGTAGTCCAGGCAATGATTTTCAAAATAAC 297
D 143 -----ProAspProHisSerArg-----LysSerHis 151
QY 298 TTAATATCAGGATTTTACGAGATATGTTACTCTATGAGCGACCAAGAAATAGAACTT 357
D 152 LeuIleSerSerPheSer---AspLeuGlyValThrLeuAspPheProSerSerAsnLeu 170
QY 358 GAAGCTGATATCAACATTTAATCCAAAAACACGATACATACATGATGATGATGAT 417
D 171 ArgPhePheLeuValArgGlySerProPheIleThrPheSerValAsnSerSerIleThr 190
QY 418 GAATACTATAACATTTTGCATATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTA 477
D 191 IleSerThrIleHisAlaValLeuSerLeuSerGlyAsnThrSerSerThrIleThr 210
QY 478 GTACTTAAAAATGACGGCATACTTTTATGTCATGATGGTTAATPACTTGCTATGACATT 537
D 211 ValLysLeuAsnAsnAsnGlnThrTrpLeu---IleTyrAlaSerSerProIleAsnLeu 229
QY 538 ACAGCTGAAGGATATCTTCGTA----- 561
D 230 ThrLysAspGlyValSerSerIleAsnCysGlyAspGlyPheSerGlyIleIleArgIle 249
QY 562 -----CCATAT-----GCATGTGCA 576
D 250 ValValLeuProAsnProAsnProTyrPheGluThrIleLeuAspGlyPheSerCysSer 269
QY 576 ----- 576
D 270 TyrProValSerGlyAspAlaAspPheThrLysProPheAlaLeuGluTyrLysTrpGlu 289
QY 577 -----GGTATAGGACGATCTTACTACTATTTTAAAGACCTCAATCTAAA----- 624
D 290 LysArgGlyTyrGly---AspLeuLeuMetLeuAlaHisProLeuHisLeuLysLeuLeu 308
QY 625 -----TTTGCTTACCAA----- 636
D 309 SerThrAsnAspCysSerIleThrValLeuAspAsnPheLysTyrAsnSerIleAspGly 328
QY 637 -----GGAAAAATAGGTATTAGTACCCCTATCACACCAAGATCTCTGCAATTATT 687
D 329 AspLeuValGlyValIleGlyAspSerTrpValLeuLysProAspPro-----Val 345

```

```

QY 688 GGTGATACCTACCATGGCGTTATTGGTAATAAATTTGAGAAGATACCTGTTAATCTCT 747
D 346 SerValThrTrpHisSerIleGlyValGlnGluAspSerHisGlnGluIleSer 365
QY 748 GTAGTATTAAATGATGCTCTCAAAACACACATCTCTCAGTAACCTCTTGACGTTGGATAC 807
D 366 AlaLeuIleLysAspValAsnAlaLeuAspSerSerAlaGluValThrAsnSerSerTyr 385
QY 808 TTT 810
D 386 Phe 386

```

## RESULT 29

B23496

TyB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1-17

C:Species: Saccharomyces cerevisiae

C:Date: 20-Aug-1987 #sequence\_revision 20-Aug-1987 #text\_change 19-Apr-2002

C:Accession: B23496

R:Warrington, J.R.; Waring, R.B.; Newlon, C.S.; Indge, K.J.; Oliver, S.G.

Nucleic Acids Res. 13, 6679-6693, 1985

A:Title: Nucleotide sequence characterization of Ty 1-17, a class II transposon from yeast

A:Reference number: A93591; MUID:86041864; PMID:2997719

A:Accession: B23496

A:Molecule type: DNA

A:Residues: 1-1348 <VAR>

A:Cross-references: EMBL:X03840; EMBL:J01333; EMBL:X06033; EMBL:X02991

C:Genetics:

A:Cross-references: SGD:S0000524

A:Mobile element: retrotransposon Ty1-17

C:Superfamily: TyB protein

Alignment Scores:

Pred. No.: 2.43 Length: 1348  
Score: 93.00 Matches: 66  
Percent Similarity: 36.60% Conservative: 31  
Best Local Similarity: 24.91% Mismatches: 93  
Query Match: 6.28% Indels: 75  
DB: 2 Gaps: 15

US-10-062-624-41 (1-840) x B23496 (1-1348)

```

QY 21 AGTAAGAGCGGTAAATCTCATTAAATGCTCAATCTTACCATATCAGTCTTTTGCA----- 75
D 802 SerLysLysArgIleAsnLeuIle-AlaAlaIleGlyValLysSerIleLysProVa 821
QY 76 -----GATCTGTAGTTCAGAACATAATGATACAAAGAA---GG 113
D 821 LArgThrThrLeuArgTyrAspGluAlaIleThrTyrAsnLysAspAsnLysGluLysAs 841
QY 114 CTTCTCATATGTCGAAAGTACATCCAGTATATATCAGTCTTACAAATTTCTGCTGTA 173
D 841 pArgTyrValGluAla---TyrHisLysGluIleSerGlnLeuLeuLysMetAsnThrTr 860
QY 174 AGAAACT-----CCTATTAAATGGAACAAATTTCTCT 203
D 860 pAspThrAsnLysTyrTyrAspArgAsnAspIleAspProLysLysValIleAsnSerMe 880
QY 204 CACTAAAAAGTTTTTCGGACTAAAGAAAGAT----- 234
D 880 tPhe-----IlePheAsnLysLysArgAspGlyThrHisLysAlaArgPheValAlaAr 898
QY 235 -GGTGATATACAAAAAGACGATTTTACAAGAGTAGCTCCAGGCATTGATTTTCAAAA 293
D 898 gGlyAspIleGlnHisProAspThrTyrAspSer-----AspMetGlnSe 913
QY 294 TAACTTAATATCAGATTTTCA-----GGAAGTATTGGTTACTCTATGGACGGA----- 342
D 913 rAsnThrValHisIstYrAlaLeuMetThrSerLeuSerIleAlaLeuAspAsnAspTy 933
QY 343 -----CCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCAAAAAACACCGA 395
D 933 rTyrIleThrGlnLeuAspIleSerSerAlaTyrLeuTyrAlaAspIleLysGlu----- 951

```

1302	tPhe-----IlePheAsnLysLysArgAspGlyThrHisLysAlaArgPheValAlaAr	1320	
QY	238	---GATATAACAAACAAAGACGATTTTCAAGAGTAGCTCCAGGCGATTGATTTTCAAAA	293
Db	1320	gSerAspIleGlnHisProAspThrTyRAspSer-----AspMetGlnSe	1335
QY	294	TAACTTTAATATCATCAGGATTTTCA-----GGAAGTATTGGTTACTCTATGCACCGA-----	342
Db	1335	rAsnThrValHisHisTyRAlaLeuMetThrSerLeuSerIleAlaLeuAspAsnAspTy	1355
QY	343	-----CCAAGNATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGA	395
Db	1355	rTyRileThrGlnLeuAspIleSerSerAlaTyRLeuTyRAlaAspIleLysGlu-----	1373
QY	396	TAACAATGATACTGATAATGGTGAATACTAT-----AAACATTTTTCATTATC	443
Db	1374	-----GluLeuTyRileArgProProHisLeuGlyLeuAs	1386
QY	444	TCGTAAAGATGCNAATGGAGATAGACGACAATATGTAGTACTTTAAAAATGACGGCATA-----	498
Db	1386	nAspLysLeuLeuArgLeuArgLysSerLeuTyRGlyLeuLysGlnSerGlyAlaAsnTr	1406
QY	499	-----ACTTTTATCTCATGATGGTTAATACTTCTGCTATGACATTTACAGCTGAAGGAGT	551
Db	1406	pTyRGluThrIleLysSerTyRLeuIleAsnCySAspMetGlnGluValArgGlyTr	1426
QY	552	ATCTTTTCGTA-----CCATATGCATGTCAGGTATAGGAGCAGCATCTTAT	596
Db	1426	pSerCysValPheLysAsnSerGlnValThrIleCysLeuPheValAsp---AspMetIl	1445
QY	597	CACATATTTTAAAGACCTCAAT-----CTAAAAATTCGCTTA	632
Db	1445	eLeuPheSerLysAspLeuAsnAlaAsnLysLysIleIleThrThrLeuLysLysGlnTy	1465
QY	633	CCAAGGAAAAATA	645
Db	1465	rAspThrLysIle	1469
RESULT 31			
S69953			
TyB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty2.DR			
N;Alternate names: protein YD9320A.14; protein YD_A_13_A			
C;Species: Saccharomyces cerevisiae			
C;Date: 09-Mar-1996 #sequence revision 06-Sep-1996 #text_change 20-Jun-2000			
C;Accession: S69953; S67321; S70226			
R;Murphy, L.; Harris, D.			
submitted to the EMBL Data Library, December 1995			
A;Reference number: S6117			
A;Accession: S69953			
A;Molecule type: DNA			
A;Residues: 1-1709 <MUR>			
A;Cross-references: EMBL:Z68329			
A;Note: biosynthesis of this protein involves a +1 frameshift in the codon for			
A;Accession: S67321			
A;Molecule type: DNA			
A;Residues: 'MTN', 429, 'V', 431-1709 <MUR>			
A;Cross-references: EMBL:Z68329; NID:G1136205; PIDN:CAA92721.1; PID:G1136219			
A;Note: the difference at the amino end is due to translation from an incorrect			
R;Wambutt, R.; Wedler, H.; Wedler, E.; Scharfe, M.			
submitted to the Protein Sequence Database, July 1996			
A;Reference number: S67608			
A;Accession: S70226			
A;Molecule type: DNA			
A;Residues: 1693-1770 <WAM>			
A;Cross-references: EMBL:Z74387			
C;Genetics:			
A;Map position: 4R			
A;Mobile element: retrotransposon Ty2.DR			
C;Superfamily: TyB protein			
Alignment Scores:			
Pred. No.:	2	5	Length:
			1770

Score:	93.00	Matches:	66
Percent Similarity:	36.60%	Conservative:	31
Best Local Similarity:	24.91%	Mismatches:	93
Query Match:	6.28%	Indels:	75
DB:	2	Gaps:	15

US-10-062-624-41 (1-840) x S69953 (1-1770)

Qy	21	AGTAAGAAGCGGTTAATCTCAATTAATGTGTCAAATCTTACCATATACAGTCTTTTTCGA	-----	75
Db	1224	SetLysLysArgIleAenLeuIleAlaAlaIleLysGlyValLysSerIleLysProVa	1243	
Qy	76	-----CATCCTGTAGGTTCAAGAACTAATGATCAACCAAGAA	---GG	113
Db	1243	lArgThrThrLeuArgTyrAspGluAlaIleThrTyrAsnLysAspAsnLysGluLysAs	1263	
Qy	114	CTTCTCATATTAGTCAAAAGTACAACATCAACAGTATATACACCTTTAGAAAAATCTCTGCTGA	173	
Db	1263	pArgTyrValGluAla--TyrHisLysGluIleSerGlnLeuLysMetAsnThrTr	1282	
Qy	174	AGAAACT-----CCTATTAAATGGAACAATCTCTCT	203	
Db	1282	pAspThrAsnLysTyrTyrAspArgAenAspIleAspProLysLysValIleAsnSerMe	1302	
Qy	204	CACATAAAAAAGTTTTCCGACTAAAGAAGAT-----	234	
Db	1302	tPhe-----IlePheAsnLysLysArgAspGlyThrHisLysAlaArgPheValAlaAr	1320	
Qy	235	-GGTGATATAACAAAAAGACGATTTACAAGAGTAGTCTCCAGGCATATGATTTTCAAAA	293	
Db	1320	gGlyAspIleGlnHisProAspThrTyrAspSer-----AspMetGlnSe	1335	
Qy	294	TAACCTTAATATCAGGATTTTCA-----GGAAGTATTGGTTACTCTATGACCGCA	342	
Db	1335	rAsnThrValHisTyrAlaLeuMetThrSerLeuSerIleAlaLeuAspAenAspTy	1355	
Qy	343	-----CCAAGAATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGA	395	
Db	1355	rTyIleThrGlnLeuAspIleSerSerAlaTyrLeuTyrAlaAspIleLysGlu	1373	
Qy	396	TAACAATGATCTACTGAATGGTGAATCTACT-----AAACATTTTGCATTTATC	443	
Db	1374	-----GluLeuTyrIleArgProProHisLeuGlyLeuAs	1386	
Qy	444	TCGTAAAGATGCAATGGAAGATCAGCAATATGATAGTACTTTAAAAATGACCGCAT	498	
Db	1386	nAspLysLeuLeuArgLeuArgLysSerLeuTyrGlyLeuLysGlnSerGlyAlaAsnTr	1406	
Qy	499	-----ACTTTTATGTCATGTGATGTTAACTACTTGTCTATGACATTTACAGCTGAAGAGT	551	
Db	1406	pTyrGluThrIleLysSerTyrIleuIleAsnCysAspMetGlnGluValArgGlyTr	1426	
Qy	552	ATCTTTTCGTA-----CATATGCAATGTGAGGTATAGGACGACATCTTAT	596	
Db	1426	pSerCysValPheLysAsnSerGlnValThrIleCysLeuPheValAsp---AspMetIl	1445	
Qy	597	CACATATTTTAAAGACCTTCAAT-----CTAAAAATTTGCTTA	632	
Db	1445	eLeuPheSerLysAspLeuAsnAlaAsnLysIleIleThrThrLeuLysLysGlnTy	1465	
Qy	633	CCAAGGAAAAATA	645	
Db	1465	rAspThrLysIle	1469	

RESULT 32

S70233  
TyB protein - yeast (*Saccharomyces cerevisiae*) retrotransposon Ty2.0  
N;Alternate names: protein O6304  
C;Species: *Saccharomyces cerevisiae*.  
C;date: 09-Mar-1996 #sequence\_revision 06-Sep-1996 #text\_change 20-Jun-2000  
C;Accession: S70233  
R;Goffeau, A.; Purnelle, B.  
Submitted to the Protein Sequence Database, July 1996



```
QY 633 CCAAGGAAAAATA 645
Db 1465 rAspThrLysIle 1469

RESULT 33
TyB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty2.0
N:Alternate names: protein O4785
C:Species: Saccharomyces cerevisiae
C>Date: 09-Mar-1996 #sequence_revision 06-Sep-1996 #text_change 20-Jun-2000
C:Accession: S70230
R:Hughes, B.; Pohl, T.M.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66685
A:Accession: S70230
A:Molecule type: DNA
A:Residues: 1-1770 <HUG>
A:Cross-references: EMBL:Z75100; NID:g1420458; PIDN:CAA99402.1; PID:g1420460
A:Note: biosynthesis of this protein involves a +1 frameshift in the codon for residue 4
C:Genetics:
A:Map position: 15R
A:Mobile element: retrotransposon Ty2.0
C:Superfamily: TyB protein

Alignment Scores:
Pred. No.: 2.5 Length: 1770
Score: 93.00 Matches: 66
Percent Similarity: 36.60% Conservative: 31
Best Local Similarity: 24.91% Mismatches: 93
Query Match: 6.28% Indels: 75
DB: 2 Gaps: 15

US-10-062-624-41 (1-840) x S70230 (1-1770)
QY 21 AGTAAGAGCGGTTAAATCTCAATGCTCAATCTTACCATATCAGTCTTTTGCA----- 75
Db 1224 SerLysLysArgIleAsnLeuIle-AlaAlaIleLysGlyValLysSerIleLysProVa 1243
QY 76 -----GATCCTGTAGTTCAGAACTAATGATACAAAGAA---GG 113
Db 1243 largThrThrLeuArgTyrAspGluAlaIleThrTyrAsnLysAspAsnLysGluLysAs 1263
QY 114 CTTCTACATTAGTCAAGATCAATCCAGATATACACACTTTAGAAATTTCTCTGCTGA 173
Db 1263 pArgTyrValGluAla---TyrHisLysGluIleSerGlnLeuLysMetAsnThrTr 1282
QY 174 AGAAACT-----CCTATTATGGAACAAATTTCTCT 203
Db 1282 pAspThrAsnLysTyrTyrAspArgAsnAspIleAspProLysLysValIleAsnSerMe 1302
QY 204 CACTAAAAAGCTTTTCGGACTAAAGAAAGAT----- 234
Db 1302 tPhe-----IlePheAsnLysLysArgAspGlyThrHisLysAlaArgPheValAlaAr 1320
QY 235 -GGTGATATACAAAAAGAGATTTTACAAGAGTAGCTCCAGCATTTGATTTTCAAAA 293
Db 1320 gGlyAspIleGlnHisProAspThrTyrAspSer-----AspMetGlnSe 1335
QY 294 TAACCTAATATACAGATTTTCA-----GGAAGTATTGGTTACTATGGACGGA----- 342
Db 1335 rAsnThrValHisHisTyrAlaLeuMetThrSerLeuSerIleAlaLeuAspAsnAspTy 1355
QY 343 -----CCAAGAATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGA 395
Db 1355 rTyrIleThrGlnLeuAspIleSerSerAlaTyrLeuTyrAlaAspIleLysGlu----- 1373
QY 396 TAACATGATATCTGATATGTTGATATCTAT-----AAACATTTTGCATATC 443
Db 1374 -----GluLeuTyrIleArgProProHisLeuLysGluLeuAs 1386
QY 444 TCGTAAGATCAATGGAGATCAGCATATGTAGTACTTTAAATGACGCATA----- 498
Db 1386 nAspLysLeuLeuArgLeuArgLysSerLeuTyrGlyLeuLysGlnSerGlyAlaAsnTr 1406
```

```
QY 499 -----ACTTTTATGTCATTGATGGTTAATACTTCTGCTATGACATTACAGCTGAAGGAGT 551
Db 1406 pTyrGluThrIleLysSerTyrLeuIleAsnCysAspMetGlnGluValArgGlyTr 1426
QY 552 ATCTTTTCGTA-----CCATATGCATGTGCAGGTATAGAGCAGATCTTAT 596
Db 1426 pSerCysValPheLysAsnSerGlnValThrIleCysLeuPheValAsp---AspMetIle 1445
QY 597 CACTATTTTTAAACACCTCAAT-----CTAAAATTTGCTTA 632
Db 1445 eLeuPheSerLysAspLeuAsnAlaAsnLysLysIleIleThrThrLeuLysGlnTy 1465
QY 633 CCAAGGAAAAATA 645
Db 1465 rAspThrLysIle 1469

RESULT 34
S69966
TyB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty2.L
N:Alternate names: protein L9931.7 b
C:Species: Saccharomyces cerevisiae
C>Date: 09-Mar-1996 #sequence_revision 06-Sep-1996 #text_change 06-Sep-1996
C:Accession: S69966
R:Favell, A.
submitted to the EMBL Data Library, January 1995
A:Description: The sequence of S. cerevisiae cosmid 9931.
A:Reference number: S59376
A:Accession: S69966
A:Molecule type: DNA
A:Residues: 1-1770 <FAV>
A:Cross-references: EMBL:U20162
A:Note: biosynthesis of this protein involves a +1 frameshift in the codon for residue 4
C:Genetics:
A:Map position: 12R
A:Mobile element: retrotransposon Ty2.L
C:Superfamily: TyB protein

Alignment Scores:
Pred. No.: 2.5 Length: 1770
Score: 93.00 Matches: 66
Percent Similarity: 36.60% Conservative: 31
Best Local Similarity: 24.91% Mismatches: 93
Query Match: 6.28% Indels: 75
DB: 2 Gaps: 15

US-10-062-624-41 (1-840) x S69966 (1-1770)
QY 21 AGTAAGAGCGGTTAAATCTCAATGCTCAATCTTACCATATCAGTCTTTTGCA----- 75
Db 1224 SerLysLysArgIleAsnLeuIle-AlaAlaIleLysGlyValLysSerIleLysProVa 1243
QY 76 -----GATCCTGTAGTTCAGAACTAATGATACAAAGAA---GG 113
Db 1243 largThrThrLeuArgTyrAspGluAlaIleThrTyrAsnLysAspAsnLysGluLysAs 1263
QY 114 CTTCTACATTAGTCAAGATCAATCCAGATATACACACTTTAGAAATTTCTCTGCTGA 173
Db 1263 pArgTyrValGluAla---TyrHisLysGluIleSerGlnLeuLysMetAsnThrTr 1282
QY 174 AGAAACT-----CCTATTATGGAACAAATTTCTCT 203
Db 1282 pAspThrAsnLysTyrTyrAspArgAsnAspIleAspProLysLysValIleAsnSerMe 1302
QY 204 CACTAAAAAGCTTTTCGGACTAAAGAAAGAT----- 234
Db 1302 tPhe-----IlePheAsnLysLysArgAspGlyThrHisLysAlaArgPheValAlaAr 1320
QY 235 -GGTGATATACAAAAAGAGATTTTACAAGAGTAGCTCCAGCATTTGATTTTCAAAA 293
Db 1320 gGlyAspIleGlnHisProAspThrTyrAspSer-----AspMetGlnSe 1335
QY 294 TAACCTAATATACAGATTTTCA-----GGAAGTATTGGTTACTATGGACGGA----- 342
```



A;Reference number: S64071  
A;Accession: S69844  
A;Molecule type: DNA  
A;Residues: 1-1770 <R1E>  
A;Cross-references: EMBL:Z72946; NID:g1323273; PIDN:CAA97177.1; PID:g1323276  
A;Genetics: CH7  
A;Note: biosynthesis of this protein involves a +1 frameshift in the codon for residue 4  
R;Naïtoui, M.; Ozawa, M.; Saeenuma, S.; Kobayashi, M.; Hagiwara, H.; Shibata, T.; Hanaoka  
Yeast 12, 77-84, 1996  
A;Title: Sequencing of a 23 kb fragment from Saccharomyces cerevisiae chromosome VI.  
A;Reference number: S61731; MUID:96381249; PMID:8789262  
A;Accession: S61736  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 'MTTN', 429, 'V', 431-1770 <NAI>  
A;Cross-references: EMBL:D44604  
A;Genetics: CH6  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1994  
C;Genetics: <CH6>  
A;Map position: 6L  
A;Mobile element: retrotransposon Ty2.F  
A;Note: YFL002w-a  
C;Genetics: <CH7>  
A;Map position: 7R  
A;Mobile element: retrotransposon Ty2.GR  
C;Superfamily: TyB protein

Alignment Scores:  
Pred. No.: 2.5 Length: 1770  
Score: 93.00 Matches: 66  
Percent Similarity: 36.60% Conservative: 31  
Best Local Similarity: 24.91% Mismatches: 93  
Query Match: 6.28% Indels: 75  
DB: 2 Gaps: 15

US-10-062-624-41 (1-840) x S58651 (1-1770)

QY 21 AGTAAGAGCGGTTAATCTCATTAATGTCATCTTACCATATCAGTCTTTTCCA----- 75  
Db 1224 SerLysLysArgIleAenLeuIle-AlaAlaIleLysGlyValLysSerIleLysProVa 1243  
QY 76 -----GATCTGTAGGTTCCAGAACTAATGATAACAAACAA-----GG 113  
Db 1243 LaArgThrLeuArgTyRaspGluAlaIleThrTyRAsnLysAspAsnLysGluLysAs 1263  
QY 114 CTCTCATATTAGTGCAGAAAGTACAAATCAAGTATATCACACTTTAGAAAAATTCCTCTCTGA 173  
Db 1263 paGtyTyValGluAla---TyrHisLysGluIleSerGlnLeuLysMetAsnThrTr 1282  
QY 174 AGAAACT-----CCTATTAAATGGACAAATTTCTCT 203  
Db 1282 paSpThrAsnLysTyTyTyRaspArgAsnAspIleAspProLysLysValIleAsnSerMe 1302  
QY 204 CATATAAAAAATTTTCGGACTAAAGAAAGAT----- 234  
Db 1302 tPhe-----IlePheAsnLysLysArgAspGlyThrHisLysAlaArgPheValAlaAr 1320  
QY 235 -GGTGATATAACAAAAAGACGATTTTACAGAGTAGTCCAGGCATTGATTTTCAAAA 293  
Db 1320 gGlyAspIleGlnHisProAspThrTyRaspSer-----AspMetGlnSe 1335  
QY 294 TAACCTTAATATCAGGATTTTCA-----GGAAGTATGTTTACTCTATGACCGA----- 342  
Db 1335 rAsnThrValIHisItyrAlaLeuMetThrSerLeuSerIleAlaLeuAspAsnAspTy 1355  
QY 343 -----CCAAGATAGAACTTGAAGTCGATCAACAAATTTAAATCCAAAAACACCGA 395  
Db 1355 rTyIleThrGlnLeuAspIleSerSerAlaTyLeuTyRAlaAspIleLysGlu----- 1373  
QY 396 TAACAATGATCTGATTAATGGTGAATACTAT-----AAACATTTTCGATTATC 443  
Db 1374 -----GluLeuTyIleArgProProHisLeuGlyLeuAs 1386

Db 1303 tPhe-----IlePheAsnLysLysArgAspGlyThrHisLysAlaArgPheValAlaAr 1321  
Qy 235 -GGTGATTAACAAAAAGACGATTTCACAAGAGTAGCTCCAGGCATTGATTTTCAAAA 293  
Db 1321 gGlyAspIleGlnHisProaspThrTyraSpSer-----AspMetGlnSe 1336  
Qy 294 TAACTTAATAATCAGGATTTTCA-----GGAAGTATTGGTTACTCTATGACCGA----- 342  
Db 1336 xAsnThrValHisHisTyraLeuMetThrSerLeuSerIleAlaLeuAspAsnAspTy 1356  
Qy 343 -----CCAAGAATAGAACCTTGAAGCTGCATCAACAATTTAATCCAAAAACACCGA 395  
Db 1356 rTyrIleThrGlnLeuAspIleSerSerAlaTyLeuTyraLaAspIleLysGlu----- 1374  
Qy 396 TAACAATGATAGTATAATGGTGAATACTAT-----AAACATTTTGCATTATC 443  
Db 1375 -----GluLeuTyriLeArgProProHisLeuGlyLeuAs 1387  
Qy 444 TCGTAAAGATGCAATGGAAGATCACCAATATGTAGTACTTAAANAATCACGCCATA----- 498  
Db 1387 nAspLysLeuLeuArgLeuArgLysSerLeuTyrgLyLeuLysGlnSerGlyAlaAsnTr 1407  
Qy 499 -----ACITTTATGTCATGTATGGTTAATACTTGTCTATGACATTACAGCTGAAGAGT 551  
Db 1407 pTyrgLuthrIleLysSerTyrlleuIleAsnCysCysAspMetGlnGluValArgGlyTr 1427  
Qy 552 ATCTTTTCGTA-----CCATATGTCATGTGCAGGTATAGGACGACAGATCTTAT 596  
Db 1427 pSerCyseValPhelyshsnserGlnValThrilcysLeuphevalasp---AspMetIl 1446  
Qy 597 CACTATTTTTAAAGACCTCAAT-----CTAAAAATTTGCTTA 632  
Db 1446 eLeuPheSerLysAspLeuAenAlaAsnLysLysilethrThrLeuLysLysGlnTy 1466  
Qy 633 CCAAGGAAAATA 645  
Db 1466 rAspThrLysile 1470

RESULT 38  
S69973  
TyB protein - yeast (*Saccharomyces cerevisiae*) retrotransposon Ty2.N  
N;Alternate names: protein N2715  
C;Species: *Saccharomyces cerevisiae*  
C;Date: 09-Mar-1996 #sequence\_revision 06-Sep-1996 #text\_change 20-Jun-2000  
C;Accession: S69973  
R;Duesterhoeft, A.; Floeth, M.; Fritze, C.; Heuss-Neitzel, D.; Hilbert, H.; Moestl, D.  
submitted to the Protein Sequence Database, April 1996  
A;Reference number: S62944  
A;Accession: S69973  
A;Molecule type: DNA  
A;Residues: 1-1810 <EUB>  
A;Cross-references: EMBL:Z71312; NID:g1301881; PIDN:CAA95902.1; PID:g1301883  
A;Note: biosynthesis of this protein involves a +1 frameshift in the codon for residue 4  
C;Genetics:  
A;Map position: 14L  
A;Mobile element: retrotransposon Ty2.N  
C;Superfamily: TYB protein

QY	76	-----GATCTGTAGGTTCAAGAACTAATGATACAAAGAA-----GG	113
DB	1283	lArgThrThrLeuArgTyrAspGluAlaIleIleTyrAsnLysAspAsnLysGluLysAs	1303
QY	114	CTTCTACATAGTCAAAAGTACATCAAGTATATACACATTTAGAAAAATCTCTGCTGA	173
DB	1303	pArgTyrValGluAla---TyrHisLysGluIleSerGlnLeuLeuLysMetAsnThrTr	1322
QY	174	AGAAACT-----CCTATTAAATGGAACAAATTTCTCT	203
DB	1322	pAspThrAsnLysTyrTyrAspArgAsnAspIleAspProLysLysValIleAsnSerMe	1342
QY	204	CACATAAAAAGTTTTCGACGACTAAAGAAGAT-----	234
DB	1342	tPhe-----IlePheAsnLysLysArgAspGlyThrHisLysAlaArgPheValAlaAr	1360
QY	235	-GGTGATATAACAAAAAAGACGATTTTACAAGAGTAGCTCCAGGCATTGATTTTCAAAA	293
DB	1360	gGlyAspIleGlnHisProAspThrTyrAspSer-----AspMetGlnSe	1375
QY	294	TAACCTTAATACAGGATTTTCA-----GGAAGTATTGGTTACTCTATGACGCGA	342
DB	1375	rAsnThrValHisHisTyrAlaLeuMetThrSerLeuSerIleAlaLeuAspAsnAspTy	1395
QY	343	-----CCAAGATAGAACCTTGAAGCTGCATATCAACAATTTTAATCCAAAAACACCGA	395
DB	1395	rTyrIleThrGlnLeuAspIleSerSerAlaTyrLeuTyrAlaAspIleLysGlu-----	1413
QY	396	TAACAATGATACTAATATGGTGAATACTAT-----AAACATTTTGCAATTATC	443
DB	1414	-----GluLeuTyrIleArgProProHisLeuGlyLeuAs	1426
QY	444	TCGTAAAGATGCAATGGAAGATCAGCAATATGTAGTACTTTAAAAATCAGCGCATA	498
DB	1426	nAspLysLeuLeuArgLeuArgLysSerLeuTyrGlyLeuLysGlnSerGlyAlaAsnTr	1446
QY	499	-----ACTTTTATGTCATTGATGGTTAACTACTTGTCTATGACATTTACAGCTGAAGGAGT	551
DB	1446	pTyrGluThrIleLysSerTyrLeuIleAsnCysAspMetGlnGluValArgGlyTr	1466
QY	552	ATCTTTTCGTA-----CCATATGCATGTCAGGTATAGCAGCAGATCTTAT	596
DB	1466	pSerCysValPheLysAsnSerGlnValThrIleCysLeuPheValAsp---AspMetI	1485
QY	597	CACATATTTTAAAGACCTCAAT-----CTAAAAATTTGCTTA	632
DB	1485	eLeuPheSerLysAspLeuAsnAlaAsnLysLysIleIleThrThrLeuLysLysGlnTy	1505
QY	633	CCAAGGAAAAATA	645
DB	1505	rAspThrLysIle	1509
RESULT 39			
C37440			
hypoetical protein AGR_C1196 [imported] - Agrobacterium tumefaciens (strain			
C.Species: Agrobacterium tumefaciens			
C.Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002			
R.Accession: C37440			
R.Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.;			
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Mar			
Science 294, 2323-2328, 2001			
A.Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobact			
A.Reference number: A97359; PMID:11743194			
A.Accession: C37440			
A.Status: preliminary			
A.Molecule type: DNA			
A.Residues: 1-219 <KUR>			
A.Cross-references: GB:AE007869; PIDN:AAK86476.1; PID:g15155624; GSPDB:GN00169			
C.Genetics:			
A.Gene: AGR_C1196			
A.Map position: circular chromosome			
Alignment Scores:			

```
Pred. No.: 2.24 Length: 219
Score: 92.50 Matches: 38
Percent Similarity: 36.26% Conservative: 24
Best Local Similarity: 22.22% Mismatches: 60
Query Match: 6.25% Indels: 49
DB: 2 Gaps: 7

US-10-062-624-41 (1-840) x C97440 (1-219)

Qy 238 GATATAACAAAAAGAC-----GATTTTACAAGAGTAGCTCCAGGCATTGATTTTCAA 291
Db 41 AspValThrValSerAspGlyProAspPheArgAlaGlyTrpGluGlyLysSerPheSer 60
Qy 292 AATAACTTAATATACAGATTTTCAGGA-----AGT 321
Db 61 AsnProProTyrTrpGlyValArgGlyThrTyrTrpPheAspGluGlyGlnLeuArgAsn 80
Qy 322 ATTGGTTACTCTATGGACGACCAAGATGAACCTTGAAGCTGCATATCAACAATTTAAT 381
Db 81 PheGlyLysLeuAspPheThrHisAspLysValTyrAla-----94
Qy 382 CCAAAAACCCGATAACAATGATACATAATGGTGAATACTATAAACAATTTTGCATTA 441
Db 95 -----AspAspGluThrLeuAlaArgSerGlyTrpSerHisPheGluPhe 109
Qy 442 TCTCGTAAAGATCAATGGAAGATCAGCAATATAGTACTTAAAAATGACGGCATAACT 501
Db 110 Thr-----AspGlyLeuAsn 114
Qy 502 TTTATGTCATTGATGGTTAATACTTGTCTATCAGCATTACAGCTGAAGAGTATCTTTCGTA 561
Db 115 LeuLeuThrLeu-----AsnAlaLeuTyrArgPheProLeuGluSerLeuProIleThr 132
Qy 562 CCATATGCATGTGCAGGTATAGAGCAGATCTT-----ATCAGTATTTTAAAGACCTC 615
Db 133 ProTyrValGlyAlaGlyValGlyIleAsnValProHisValGluValTyrArgProSer 152
Qy 616 AATCTAAATTTGCTTAC-----CAAGGAAAAATAGGTATTAGT 654
Db 153 GlyLysThrPheGluTyrGlnPheGlyAlaThrLeuGlnAlaGlnAlaGlyLeuSer 172
Qy 655 TACCTATACACACAGAGTCTCTGCATTTATT 687
Db 173 TyrArgIleThrAspSerTrpSerThrPheVal 183

RESULT 40
AF2658
conserved hypothetical protein Atu0668 [imported] - Agrobacterium tumefaciens (strain C5
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C/Accession: AF2658
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, F.; Zhang, S.
Science 294, 2317-2323, 2001
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58..
A/Reference number: AB2577; PMID:11743193
A/Accession: AF2658
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-219 <KUR>
A/Cross-references: GB:AE008688; PID:AL41684.1; PID:g17739029; GSPDB:GN00186
A/Experimental source: strain C58 (Dupont)
C/Genetics:
A/Gene: Atu0668
A/Map position: circular chromosome

Alignment Scores:
Pred. No.: 2.24 Length: 219
Score: 92.50 Matches: 38
Percent Similarity: 36.26% Conservative: 24
```

```
Best Local Similarity: 22.22% Mismatches: 60
Query Match: 6.25% Indels: 49
DB: 2 Gaps: 7

US-10-062-624-41 (1-840) x AF2658 (1-219)

Qy 238 GATATAACAAAAAGAC-----GATTTTACAAGAGTAGCTCCAGGCATTGATTTTCAA 291
Db 41 AspValThrValSerAspGlyProAspPheArgAlaGlyTrpGluGlyLysSerPheSer 60
Qy 292 AATAACTTAATATACAGATTTTCAGGA-----AGT 321
Db 61 AsnProProTyrTrpGlyValArgGlyThrTyrTrpPheAspGluGlyGlnLeuArgAsn 80
Qy 322 ATTGGTTACTCTATGGACGACCAAGATGAACCTTGAAGCTGCATATCAACAATTTAAT 381
Db 81 PheGlyLysLeuAspPheThrHisAspLysValTyrAla-----94
Qy 382 CCAAAAACCCGATAACAATGATACATAATGGTGAATACTATAAACAATTTTGCATTA 441
Db 95 -----AspAspGluThrLeuAlaArgSerGlyTrpSerHisPheGluPhe 109
Qy 442 TCTCGTAAAGATCAATGGAAGATCAGCAATATAGTACTTAAAAATGACGGCATAACT 501
Db 110 Thr-----AspGlyLeuAsn 114
Qy 502 TTTATGTCATTGATGGTTAATACTTGTCTATCAGCATTACAGCTGAAGAGTATCTTTCGTA 561
Db 115 LeuLeuThrLeu-----AsnAlaLeuTyrArgPheProLeuGluSerLeuProIleThr 132
Qy 562 CCATATGCATGTGCAGGTATAGAGCAGATCTT-----ATCAGTATTTTAAAGACCTC 615
Db 133 ProTyrValGlyAlaGlyValGlyIleAsnValProHisValGluValTyrArgProSer 152
Qy 616 AATCTAAATTTGCTTAC-----CAAGGAAAAATAGGTATTAGT 654
Db 153 GlyLysThrPheGluTyrGlnPheGlyAlaThrLeuGlnAlaGlnAlaGlyLeuSer 172
Qy 655 TACCTATACACACAGAGTCTCTGCATTTATT 687
Db 173 TyrArgIleThrAspSerTrpSerThrPheVal 183

Search completed: July 8, 2003, 10:50:23
Job time : 79.5 secs
```

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 8, 2003, 10:34:21 ; Search time 20.5 Seconds  
(without alignments)  
3399.036 Million cell updates/sec

Title: US-10-062-624-41  
Perfect score: 1481  
Sequence: 1 atgaattataagaaattctt.....ttggaatgaggttcaccttc 840

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/cgn2 1/USPTO epool/US10062624/runat 08072003 093519 2001/app query.fasta.1.1031  
-DB=SwissProt 40 -QFMT=fastan -SUFFIX=rsp -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCLALIGN=200 -THR\_SCORE=500 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptp -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10062624 @CGN 1.1.26 @runat 08072003 093519 2001 -NCPU=3  
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	386.5	26.1	282	1 MSP4 ANAMA	Q07408 anaplasma m
2	102	6.9	1176	1 RPB2 DROME	P08266 drosophila
3	97.5	6.6	1174	1 RPB2 HUMAN	P30876 homo sapien
4	95	6.4	1004	1 SLPO BACER	P09333 bacillus br
5	93.5	6.3	658	1 LYTE STRPN	Q02407 streptococc
6	93	6.3	180	1 Y788 RICPR	O05978 rickettsia
7	93	6.3	608	1 YD93 METJA	Q58788 methanococc
8	93	6.3	1770	1 YCB9 HELPJ	P25384 saccharomyc
9	92.5	6.2	439	1 YHBT YEAST	P38059 lactobacill
10	92.5	6.2	532	1 RPB2 YEAST	P38745 saccharomyc
11	91.5	6.2	1224	1 RPB2 YEAST	P08518 saccharomyc
12	89.5	6.2	547	1 NUSM ASCSU	P44884 ascaris suu
13	89.5	6.0	1287	1 VAC2 HELPJ	Q48245 helicobacte
14	89.5	6.0	1288	1 VAC2 HELPJ	Q92kw5 helicobacte
15	89	6.0	741	1 HLYA VIBCH	P09545 vibrio chol
16	89	6.0	859	1 AFAC ECOLI	P53177 escherichia
17	88.5	6.0	234	1 OPAB NEIGO	Q04874 neisseria g
18	88	5.9	1894	1 ME21 SCHPO	Q13967 schizosacch

19	87.5	5.9	793	1	YHCD ECOLI	P45420 escherichia
20	87	5.9	492	1	CL31 DROME	Q9vfj0 drosophila
21	87	5.9	492	1	MURE STAAU	Q86491 staphylococ
22	87	5.9	776	1	VP4 ROTPS	P11114 porcine rot
23	86.5	6.0	414	1	CCA_BUCAL	P57169 buchnera ap
24	86.5	5.8	601	1	CYSJ_BUCAL	P57503 buchnera ap
25	86	5.8	1314	1	SWII YEAST	P09547 saccharomyc
26	85.5	5.8	418	1	YCCE ECOLI	P36661 escherichia
27	84.5	5.7	457	1	UCR1 YEAST	P07256 saccharomyc
28	84.5	5.7	1956	1	ATX1 PLAPA	Q04956 plasmodium
29	84	5.7	347	1	YF86 MYCPN	Q04956 plasmodium
30	83.5	5.6	234	1	OPAF NEIGO	P57194 mycoplasma
31	83	5.6	234	1	PEPT_CLOPE	Q04879 neisseria g
32	83	5.6	494	1	MURE STAAU	Q8xpd8 clostridium
33	83	5.6	588	1	FTSI ECOLI	Q99v74 staphylococ
34	83	5.6	776	1	VP4 ROTF6	P04286 escherichia
35	83	5.6	878	1	OSB2 HUMAN	Q07416 feline rota
36	82.5	5.6	234	1	OPAF NEIGO	Q069r2 homo sapien
37	82.5	5.6	289	1	ROCI ARATH	Q04878 neisseria g
38	82.5	5.6	1290	1	VACA HELPJ	Q52uu4 arabidopsis
39	82	5.5	427	1	PURA DICDI	P55981 helicobacte
40	82	5.7	438	1	INX7 DROME	P21900 dictyosteli
41	82	5.5	706	1	HS7S_PEA	Q9v3w6 drosophila
42	82	5.5	776	1	VP4 ROTPY	Q02028 plasm sativ
43	81.5	5.5	628	1	LJ_HUMAN	P25174 porcine rot
44	81.5	5.5	678	1	YNC7 YEAST	P50895 homo sapien
45	81	5.5	406	1	DCDA_NEIMA	P53968 saccharomyc
						Q9jwa6 neisseria m

#### ALIGNMENTS

RESULT 1  
MSP4 ANAMA  
ID MSP4 ANAMA STANDARD; .PRT; 282 AA.  
AC Q07408;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Major surface antigen 4 precursor.  
GN MSP4.  
OS Anaplasma marginale.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Anaplasmataceae; Anaplasma.  
OX NCBI\_TaxID=770;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-53.  
RX MEDLINE=94124017; PubMed=8294020;  
RA Oberle S.M., Barbet A.F.;  
RT "Derivation of the complete msp4 gene sequence of Anaplasma marginale without cloning.";  
RL Gene 136:291-294(1993).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L01987; AAC36877.1; -  
DR InterPro; IPR002566; Surface\_Ag\_msp4.  
DR Pfam; PF01617; Surface\_Ag\_2; 1.  
KW Antigen; Signal.  
FT SIGNAL 1 29  
FT CHAIN 30 282 MAJOR SURFACE ANTIGEN 4.  
SQ SEQUENCE 282 AA; 29997 MW; 30B8D31085A8834B CRC64;  
Alignment Scores:  
Pred. No.: 1.54e-25 Length: 282  
Score: 386.50 Matches: 96  
Percent Similarity: 50.34% Conservative: 54  
Best Local Similarity: 32.21% Mismatches: 113

```

Query Match: 26.10% Indels: 35
DB: 1 Gaps: 9
US-10-062-624-41 (1-840) x MSP4_ANAMA (1-282)
QY 1 ATGAATTATAAGAAATCTTA-----GTAAGAAGCGCTTAATCTCATTAATGTCA 51
DB 1 MetAsnTyrArgGluLeuPheThrGlyGlyLeuSerAlaAlaThrValCysAlaCysSer 20
QY 52 ATCTTACCATAFACGTCT---TTTCGAGATCTCT-----GTAGGTTTCAAGAACT 96
DB 21 LeuLeuValSerGlyAlaValValAlaSerProMetSerHisGluValAlaSerGluGly 40
QY 97 AATGATAACAAGAGGCTTCTACATTAGTGAAGAGTCAAT-----CCAGATATA 147
DB 41 GlyValMetGlySerPheTyrValGlyAlaAlaTyrSerProAlaPheProSerVal 60
QY 148 TCACACTTT-----AGAAAATCTCTGCTGAAGAACTCTTATTAATGGAACAATCT 201
DB 61 ThrSerPheAspMetArgGluSerSerIysGluThrSerTyrValArgGlyTyrAspIys 80
QY 202 CTCCTAAAAAGTTTTCGGACTAAGAAGATGTGTATATACAAAAAAGACGATTTT 261
DB 81 SerIleAlaThrIle-----AspValSerValProAlaAsnPhe 93
QY 262 ACAAGCTAGCTCCAGGCATTCATTTTCAAAATACTTAATATCAGGATTTTCAGAACT 321
DB 94 SerTyrSerGlyTyrThrPheAlaPheSerTyrAsnLeuIleThrSerPheAspGlyAla 113
QY 322 ATTGGTTACTCTATGACGCGCAAGATAGAACTTGAAGCTGCATATCAACAATTTAAT 381
DB 114 ValGlyTyrSerLeuGlyGlyAlaArgValGluLeuGluAlaSerTyrArgArgPhe--- 132
QY 382 CCAAAAAACACCGCATACATGATCTAGTAAATGTGTAACTACTATAAATACAT----- 432
DB 133 -----AlaThrLeuAlaAspGlyGlnTyrAlaIysSerGlyAlaGlu 146
QY 433 -----TTTGCATTATCTGTAAGATGCAATGGAAGATCAGCATATGTAGTACTTAA 486
DB 147 SerLeuAlaAlaIleThrArgAspAlaAsnIleThrGluThrAsnTyrPheValValIys 166
QY 487 AATGACGGCATAACTTTTATGTCATTGATGTTTAACTATGTCATGACATACAGCTGAA 546
DB 167 IleAspGluIleThrAsnThrSerValMetLeuAsnGlyCysTyrAspValIleHisThr 186
QY 547 GGAGTATCTTTCGTACCATCATGTCAGGTATAGGACGAGATCTTATCATCTATTTT 606
DB 187 AspLeuProValSerProTyrValCysAlaGlyIleGlyAlaSerPheValAspIleSer 206
QY 607 AAAGACCTCAATCTAAATTTGCTTACCAGGAAAAATAGGTATTAGTTACCTATCACA 666
DB 207 LysGlnValThrThrLysLeuAlaTyrArgGlyLysValGlyIleSerTyrGlnPheThr 226
QY 667 CCAGAAGTCTCTGCATTTTATGTCGTGATACCTACCATGCGGTATTGTTAAATAAATTCAG 726
DB 227 ProGluIleSerLeuValAlaGlyGlyPheTyrHisGlyLeuPheAspGluSerTyrIys 246
QY 727 AAGTACTCTGTAATAACTCTCTGTAGTATTAATAGTGTCTTCAACACCATCTCTCTCA 786
DB 247 AspIleProAlaHisAsnSerValIysPheSerGly-----GluAlaIysAlaSer 263
QY 787 GTAACCTCTGAGTTGGATCTTTTCGGCGGAGAAATTTGGAATGAGCTTCCACCTTC 840
DB 264 ValLysAlaHisIleAlaAspTyrGlyPheAsnLeuGlyAlaArgPheLeuPhe 281

```

## RESULT 2

```

RPB2 DROME STANDARD: PRT: 1176 AA..
AC P08266; Q95027; Q04155; Q9VFM7;
DT 01-AUG-1988 (Rel. 08, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerase II 140 kDa polypeptide (EC 2.7.7.6)

```

```

DE (RNA polymerase II subunit 2).
GN RPI1140 OR CG3180.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE OF 54-1176 FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=88011299; PubMed=3116266;
RT "RNA polymerase II of Drosophila Relation of its 140,000 Mr subunit
RT to the beta subunit of Escherichia coli RNA polymerase.";
RL J. Mol. Biol. 195:929-937(1987).
[2]
RP SEQUENCE OF 1-69 FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=91276237; PubMed=1905256;
RT Sizler S., Oldenburg I., Peterson G., Bautz E.K.F.;
RT "Analysis of the promoter region of the housekeeping gene DnRP140 by
RT sequence comparison of Drosophila melanogaster and Drosophila
RT virilis.";
RL Gene 100:155-162(1991).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Poser C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.D., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin G., Houston K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kinnison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.W., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -1- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM 9 TO 14
CC DIFFERENT POLYPEPTIDES. RNA POLYMERASE II CONSISTS OF 10 DIFFERENT

```









```

Db      328 PheTyrLeuLysSerAspGlyLysIleAlaGluLysGluTrpValTyrAspSerHisSer 347
Qy      265 AGAGTA-----GCTCCAGCATTTGATTTTCAAAATPACTTAATATCAGGA 309
Db      348 GlnAlaTrpTyrTyrPheLysSerGlyGlyTyrMetAlaLysAsnGluThrValAspGly 367
Qy      310 TTTTCA---GCAAGTATTGGTACTCTATGACGACCAAGATAGAACTTGAAGCTTCA 366
Db      368 TyrGlnLeuGlySerAspGlyLysTrpLeuGlyGlyLysThrTrpAsnGluAsnAlaAla 387
Qy      367 TATCAACAAATTAATCCAAAAACACCCGATACCAATGATGATAATGGTGAATACTAT 426
Db      398 TyrTyrGlnValProValThrAlaAsnValTyrAspSerAsp---GlyGluLysLeu 406
Qy      427 AAACATTTTGATATATCTCGT-----AAAGATGCAATGGAAGATCAGCAA 471
Db      407 SerTyrIleSerGlnGlySerValTrpLeuAspLysAspArgLysSerAspAspLys 426
Qy      472 TATGTAGTACTTAAATAACAGGCATACT---TTTATGTCAATGATGTTAATCTTGC 528
Db      427 ArgLeuAlaIleThrIleSerGlyLeuSerGlyTyrMetLys-----Thr 441
Qy      529 TATGACATTCAGCTGAAGAGTACTCT-----TTGCTACCATATGTCATGTGCAGGTATA 582
Db      442 GluAspLeuGlnAlaLeuAspAlaSerLysAspPheIleProTyr----- 456
Qy      583 GGAGCAGATCTTATCATTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAA 642
Db      457 -----TyrGluSerAspGlyHisArgPheTyrHisTyrValAla 469
Qy      643 ATAGGTATTAGTACCCTATCACACCAGAGTCTCTGCATT---ATTGGTGGTACTAC 699
Db      470 GlnAsnAlaSerIleProValAlaSerHisLeuSerAspMetGluValGlyLysLysTyr 489
Qy      700 CATGCGCTATTGGTAAATAATTTGAGAAGATACCTGTAATAACCTCTAGTATTAAT 759
Db      490 TyrSerAlaAspGlyLeuHisPheAspGlyPheLysLeuGluAsnProPheLeuPheLys 509
Qy      760 GATGCTCTCAAAACCATCTGCTTCA 786
Db      510 AspLeuThrGluAlaThrAsnTyrSer 518

```

## RESULT 6

```

ID Y788_RICPR STANDARD; PRT; 180 AA.
AC O05978;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein RP788.
GN RP788.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiaseae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=97419517; PubMed=9274032;
RA Andersson J.O., Andersson S.G.E.;
RT "Genomic rearrangements during evolution of the obligate
RT intracellular parasite Rickettsia prowazekii as inferred from an
RT analysis of 52015 bp nucleotide sequence.";
RL Microbiology 143:2783-2795(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sacheritz-Ponten T., Almark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of

```

```

RT mitochondria.";
RL Nature 396:133-140(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y11779; CAA72462.1; -.
KW EMBL; AJ235273; CAA15214.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 180 AA; 20473 MW; AEBEEF1F8591BF4C CRC64;

```

## Alignment Scores:

```

Pred. No.: 1.82 Length: 180
Score: 93.00 Matches: 49
Percent Similarity: 41.38% Conservative: 35
Best Local Similarity: 24.14% Mismatches: 65
Query Match: 6.28% Indels: 54
DB: 1 Gaps: 13

```

US-10-062-624-41 (1-840) x Y788\_RICPR (1-180)

```

Qy      310 TTTTCAGAGATTTGGTACTCTATGACGGACCAAGAAATAGAACTTGAA-----GCT 363
Db      2 TyrSerGlyLysIleGlyTyrIleIle---TyrProGlnIleSerMetGluPheSerAla 20
Qy      364 GCATATCAACAATTTAATCCAAAAACACCCGATACCAATGATGATGATAATGGTGAATAC 423
Db      21 ThrTyrGln-----ProLys-----Tyr 26
Qy      424 TATAAATTTTGCATTATCTCTGTAAGATGCAATGGAAGATCAGCAATAT---GTAGTA 480
Db      27 ArgLeuHisTyrSerLeuGlnHisLysAsnLeuIleAsnGlyLeuThrIleProLysThr 46
Qy      481 CTTAAATGACGGCATAACTTTTATGTCATTGATGGTTAATACTTGTCTATGACATTACA 540
Db      47 IleGlyAsnThrThrIleValSerAsnIleTyrMetLeuAsnLeuIleTyrAspLeu--- 65
Qy      541 GCTCAAGGAGTA---TCTTTGTCATCATATGCATGTGCAGGTATAGGAGCAGATCTTATC 597
Db      66 ---GluLysIleLysThrPheThrProPheIleIleLeuGlyGlyIleThrArgVal 84
Qy      598 ACTATTTTAAAGACCTCAATCTAAATTTGCT-----TACCAAGGAAAAATAGGTATTAGTACCTTATC 630
Db      85 LysVal---LysSerThrSerSerLysTrpSerLeuIleAsnAsnAspTyrPheLysVal 103
Qy      631 -----TACCAAGGAAAAATAGGTATTAGTACCTTATC 663
Db      104 HisArgThrSerLysAsnCysValThrTrpGlnAlaGlyLeuGlyIleAlaGlnHisIle 123
Qy      664 ACACCAAGTCTCT---GCAATTTATTTGGTGATACCTACCATGGCGCTTAT 711
Db      124 ThrProAspLeuSerIleAspAlaThrAlaLysLeuGlnThrAlaTyrArg-----Val 141
Qy      712 GGTAATAAATTTGAGAAGATACCTCTGTAATAACTCTCTGTAGTATTAAATGATGCTCTCAA 771
Db      142 ArgIleAsnTyrAspThrLeuAspMetLysThrValGlnLeuMetAsn----- 157
Qy      772 ACCACATCTGCTCAGTAACCTCTTGCAGCTGGATACCTTTGGCGGAGAAATTTGGAATGAGG 831
Db      158 AlaAsnSerIleLysLysThrIleSerVal-----GlyGluPheGlyIleGly 173
Qy      832 TTCACCTTC 840
Db      174 PheThrTyr 176

```

## RESULT 7

```

ID YD93_METJA STANDARD; PRT; 608 AA.

```

```

AC Q59788;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ1393.
GN MJ1393.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN RP
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96373999; PubMed=868087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-P., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RL jannaschii";
RL Science 273:1058-1073 (1996).
CC -/- SIMILARITY: TO M.JANNASCHII MJ1394 AND A.FULGIDUS AF2028.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67579; AAB99403.1; -.
DR TIGR; MJ1393; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 4 24 POTENTIAL.
SQ SEQUENCE 608 AA; 66768 MW; 010FAP1C29F8C73C CRC64;

Alignment Scores:
Pred. No.: 1.96 Length: 608
Score: 93.00 Matches: 57
Percent Similarity: 36.33% Conservative: 44
Best Local Similarity: 20.50% Mismatches: 121
Query Match: 6.28% Indels: 56
DB: 1 Gaps: 10

US-10-062-624-41 (1-840) x YD93_METJA (1-608)
QY 67 TCTTTGCGATCTGTTAGTTCAGAACTAATGATAACAAA-----GAAGGC 114
Db 204 AsnPhelLeuAnlleThrGlyAlaIleAlaAsnGluGlySerlleThrLeuTrpAspGly 223
QY 115 TTCTACATTAGTCAAGTACAACTCAAGTATATCA-----CACTTT 156
Db 224 ProTyrrPheLeuProGlyTyrrAsnAspSerLeuThrTrpThrGlyValValleAsnThr 243
QY 157 AGAAATTCCTGCTGGAAGAACTCCTATTATGCAACAATCTCTCACTAAAGATT 216
Db 244 ThrLysAsnAlaThrIleThrIleAsnIleThrGlyAsnAsnThrTyrrAsnArgThr 263
QY 217 TTCGGACTTAAAGAAGATGGTGATATAACAAAAAGACAGATTTTACAAGAGTAGTCCA 276
Db 264 GlyThrLeuMetLysTyrrGlyPheAlaVallePhePheGluPheAsnGlyThrLysSer 283
QY 277 GGCATTGATTTCAAAATACTTAATATCAGGATTTTCAGGAAGTATGGTTACTCTATG 336
Db 284 GlyThrLysIleGluGlyIleTyrrAlaThrGlyTyrr---GlyGlyValSerAlaThrLys 302
QY 337 GACGACCAAGAATA-----GACTTGAAGCTGCTATCAACAATTAAT 381
Db 337 GACGACCAAGAATA-----GACTTGAAGCTGCTATCAACAATTAAT 381

```

## RESULT 8

## YCB9\_YEAST

```

ID YCB9_YEAST STANDARD; PRT; 1770 AA.
AC P25384; P87006;
DT 01-MAY-1992 (Rel. 22, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transposon Ty2 protein B (Ty1-17 protein B).
GN TY2B OR YCLO19W OR YCLO19W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 425-1347 FROM N.A.
RX MEDLINE=86205247; PubMed=3010239;
RA Warrington J.R., Anwar R., Newlon C.S., Waring R.B., Davies R.W.,
RA Indge K.J., Oliver S.G.;
RT "A 'hot-spot' for Ty transposition on the left arm of yeast
RL Nucleic Acids Res. 14:3475-3485 (1986).
RN [2]
RP SEQUENCE OF 425-1347 FROM N.A.
RX MEDLINE=86041864; PubMed=2997719;
RA Warrington J.R., Waring R.B., Newlon C.S., Indge K.J., Oliver S.G.;
RT "Nucleotide sequence characterization of Ty 1-17, a class II
RL transposon from yeast."
RN [3]
RP SEQUENCE FROM N.A.
RX Oliver S.G., Anwar R., Brown A., Gent M.E., Indge K.J., James C.M.,
RA Staveva L.I.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RN [4]
RP REVISIONS.
RA Gromadka R.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -/- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A11.
CC -----

```



Percent Similarity: 35.12% Conservative: 38  
Best Local Similarity: 22.41% Mismatches: 125  
Query Match: 6.25% Indels: 69  
DB: 1 Gaps: 14

US-10-062-624-41 (1-840) x SLAP\_LACHE (1-439)

QY 19 CTAGTAAAGACGGTGAATCTCAATTAATGTCAATCTTACCATATCATGCTTTTCACGAT 78  
Db 7 IleValSerAlaAlaAlaAlaAlaLeuLeuAlaAlaProIleAlaAlaThrAlaMet 26  
QY 79 CCGTAGGTCACAGAACTAATGATATACAAAGAGGCTTCTAC-----ATTAGTCA 129  
Db 27 ProValAsnAlaAlaThrIleAsnAlaAspSerAlaIleAsnAlaAsnThrAsnAla 46  
QY 130 AAGTACAAT-----CCAAGTATATCACACTTTAGAAAAATCTCTGCTCAAGAA 177  
Db 47 LysTyAspValAspValThrProSerIleSerAlaIleAlaAlaValAlaLysSerAsp 66  
QY 178 ACT-----CCTATTATGAACAAATCTCTCACTAAAAAAAGTTTTCGGA--- 222  
Db 67 ThrMetProAlaIleProGlySerLeuThrGlySerIleSerAlaSerTyAsnGlyLys 86  
QY 223 -----CTAAGAAAGAT-----GGTATATTAACAAAAAAGACGAT----- 258  
Db 87 SerTyThrAlaAsnLeuProLysAspSerGlyAsnAlaThrIleThrAspSerAsnAsn 106  
QY 259 -----TTTACAAGAGTAGCTCCAGGC 279  
Db 107 AsnThrValLysProAlaGluLeuGluAlaAspLysAlaTyThrValThrValProAsp 126  
QY 280 ATTGATTTCCTCAAAATAACTTAATATATCAGGATTTTCAGGAAGTATGTTACTCTATGCAC 339  
Db 127 ValSerPhe-----AsnPheGlySerGluAsnAla 136  
QY 340 GGACCAAGATAGCACTTGAGCTGCATATCAACATTTAATTCACAAAAACACCGATAAC 399  
Db 137 GlyLysGluIleThrIleGlySerAla-----AsnProAsnValThr----- 150  
QY 400 AATGATACTGATAATCGTGAATATATAACATTTTGCATTATCTCTGAAGATGCAATG 459  
Db 151 -----PheThrGluLysThrGlyAspGlnPro 159  
QY 460 GAAGATCAGCAATATAGTACTTAAATAATGACGCGATAACT- - -TTTATGTCTATTGATG 516  
Db 160 AlaSerThrValLysValThrLeuAspGlnAspGlyValAlaLysLeuSerSerValGln 179  
QY 517 GTTAATACTTGTATGACATATACAGCTGAAGAGTATCTTCGTACCATATATGATGTCGA 576  
Db 180 IleLysAsnValTyAlaIleAspThrThrTyAsnSerAsnValAsnPheTyAspVal 199  
QY 577 GGTATAGGACGACATCTTATCTATCTATTTTAAAGACCTCACTTAAATTTGCTTACCAA 636  
Db 200 ThrThrGlyAlaThrValThrThr-----GlyAlaValSerIleAspAlaAspAsnGln 217  
QY 637 GGAATAATAGGTATTAGTTACCTATCACACAGAGTCTCTGCAATTTATTGTTGGTATAC 696  
Db 218 GlyGlnLeuAsn-----IleThrSerValValAlaAlaIleAsnSerLysTyThr 233  
QY 697 TACCATGGCGTTATTGGTAAATAA-----TTTGAGAAGATACCTGTTAATAACTCTGTAGTA 753  
Db 234 PheAlaAlaGlnTyAspLysLysGlnLeuThrAsnValThrPheAspThrGluThrAla 253  
QY 754 TTAATAGTCTCTCTCAACACCATCTGCTCAGTAACCTCTGAGCTTGGATCTTT 810  
Db 254 ValLysAspAlaLeuLysAlaGlnLysIleGluValSer---SerValGlyTyThrPhe 271

RESULT 10  
YHB7 YEAST  
ID YHB7 YEAST STANDARD; PRT; 532 AA.  
AC F38745;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)

15-JUN-2002 (Rel. 41, Last annotation update)  
Hypothetical 61.2 kDa protein in APM2-DUR3 intergenic region  
precursor.  
YHL017W.  
Saccharomyces cerevisiae (Baker's yeast).  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
NCBI\_TaxID=4932;  
(1)  
SEQUENCE FROM N.A.  
SPRAIN=5288C / AB972;  
MEDLINE=94378003; PubMed=8091229;  
Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,  
Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,  
Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,  
Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,  
Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,  
Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,  
Vaudin M.;  
"Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
VIII.";  
Science 265:2077-2082(1994).  
-1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
-1- SIMILARITY: STRONG, TO YEAST PTM1.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: U11582; AAB5070.1; -;  
PIR: S46831; S46831.  
SGD: S0001009; YHL017W.  
Hypothetical protein; Transmembrane; Signal.  
FT SIGNAL 1 20 POTENTIAL.  
FT CHAIN 21 532 HYPOTHETICAL 61.2 KDA PROTEIN IN  
FT OP11-DUR3 INTERGENIC REGION.  
FT TRANSHEM 194 214 POTENTIAL.  
FT TRANSEM 226 247 POTENTIAL.  
FT TRANSEM 263 283 POTENTIAL.  
FT TRANSEM 302 322 POTENTIAL.  
FT TRANSEM 331 351 POTENTIAL.  
FT TRANSEM 376 396 POTENTIAL.  
FT TRANSEM 416 436 POTENTIAL.  
SQ SEQUENCE 532 AA; 61205 MW; EF2CB7CB74006503 CRC64;  
Alignment Scores:  
Pred. No.: 2.15 Length: 532  
Score: 92.50 Matches: 75  
Percent Similarity: 34.91% Conservative: 58  
Best Local Similarity: 19.69% Mismatches: 99  
Query Match: 6.43% Indels: 149  
DB: 1 Gaps: 20  
US-10-062-624-41 (1-840) x YHB7 YEAST (1-532)  
QY 796 CAAGATTTACTGAAGCAGATGTGGTTGAGGAGCATCAATTAATACACAGGATTTATTA 737  
Db 116 GlnGluThrAlaIleAspProPheThrSerLysGluHisLysLeuThrSerGlnIleLeu 135  
QY 736 CAGGTATCTTCTCAAAATTTATTACCAATTAACCCCATGGTAGTATCCACCAATAATGAC 677  
Db 136 ThrPheThrGlnGln-GluLeuGlyThrAsnAspLysValTySerIleAsnLys---Th 154  
QY 676 AGACTTCTGGTGTAGAGGTAACATAACCTATTTTCCCTTGGTAAGCAATTTTAGT 617  
Db 154 rGlyTyTyTyCys---ValThrSerSerPheIleSerSerSerLysPheLysAl 173  
QY 616 TGAGGTCT---TTTAAAAATAGT-----GATAAGATCTGCTCTCTAT--- 580

Db 173 aThrValaAenPheArGAsnAlaTyRGlyGlnLeuAspAlaSerGluAlaTyRLysMetPr 193  
 QY 579 -----ACCTGCACATGCATATGTTAGCAAGATA 551  
 Db 193 oileTyRAlaPheLeuAlaValAlaTyRAlaValCysThrLeuVal-----Ty 209  
 QY 550 CTCCTTCAGCTGTAATGTCATAGCAAGATTAAACCATCAATGACAT----- 505  
 Db 209 rSerTrpLeuCys-----TrpLysHisArg-----HisGluLeuLeuProLe 223  
 QY 504 -AAAAGTTATGCGCTCAATTTT-----AA 482  
 Db 223 uGlnArgTyRileLeuValPheCysIlePheLeuThrAlaAspThrIlePheValTrpMe 243  
 QY 481 GTACTACATATTTG-----CTGAT 464  
 Db 243 tTyRTrpIleGluAsnGlnLysGlyAsnSerValAlaLeuHisValTyRMetVa 263  
 QY 463 CTTCATTCGCATCTTTACG----- 445  
 Db 263 iPheIleSerIlePheSerAlaGlyLysMetThrPheThrLeuLeuAlaLeuLeuI 283  
 QY 444 -----AGATAATGCAA 434  
 Db 283 eSerLeuGlyTyRGlyIleValTyRProLysLeuAspArgThrLeuLeuArgArgCysGl 303  
 QY 433 AATGTTTATAGTATTCACATATTCAGTATCAAT----- 400  
 Db 303 nilePheAlaIlePheThrPheAlaValCysValAlaPheLeuValGlnLysTyRSerGl 323  
 QY 399 -----GTTATCGTGTCTTTGGATAATATGTCATATGAG 362  
 Db 323 nAsnSerGluSerLeuSerAsnLeuIleThrAlaIleProLeuValCysLe 343  
 QY 361 CTTCAGTTCTATCTTGTGTCGTCAGAGTAACCAATATCTCTGAAATCTCGATA 302  
 Db 343 uPheAlaPheTyRLeuThrLeuSerSerMetAsnLysThr----- 357  
 QY 301 TTAAGTT-----ATTTGAAATCAATGCTCGAGTACTC 266  
 Db 358 -----MetThrTyRLeuArgGluGlnAsnGlnValValLysLeuAsnMetTyRArg----- 374  
 QY 265 TTGTAATATCGCTTTTGTGTATATACCATCTTTCTTTAGTCGAAACCTTTTATAG 206  
 Db 375 -----LysLeuIleLeuCysTyRileSerLeuPheIleLeuPheLeuGly-LeuLeuV 393  
 QY 205 TGAGAGAAATTT-----GTTCCATTAATAGAGATTCTTCAGCAGAGAAATTTCTAAAGT 152  
 Db 393 alSerThrPheAlaTyRValGlyMetAspThrValAspMetIleGluGlnTyRTrpLysT 413  
 QY 151 GTGATATACTT-----GGATTGTACT 131  
 Db 413 hrGluPheLeuIleThrAspThrTrpProSerPheValTyRLeuValPheValIleP 433  
 QY 130 TTGCTACTAATGAGAGCTCTTCTTGTATCATAGTTCTTCTGAA-----C 86  
 Db 433 heAlaPhePheTrpArgProThrSerTyRLeuLeuAlaCysSerHisGlnLeuP 453  
 QY 85 CTACAGGATCTCGAAAGACTGATATGTAAGATTGACATTAAATGAGATTAAACGGCCT 27  
 Db 453 roThrAspMetGluAen-----ValSerGluPheAspLeuAspIleAenSerLeu 470

RESULT 11  
 RPB2\_YEAST  
 ID RPB2\_YEAST  
 AC P08518; Q12738;  
 DT 01-AUG-1988 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE DNA-directed RNA polymerase II 140 kDa polypeptide (EC 2.7.7.6) (B150)  
 DE (RNA polymerase II subunit 2).  
 GN RPB2 OR P022 OR RPB150 OR YOR151C.  
 OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87147239; PubMed=3547406;  
 RA Sweetser D., Nonet M., Young R.A.;  
 RT "Prokaryotic and eukaryotic RNA polymerases have homologous core  
 subunits.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:1192-1196 (1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c / FV1678;  
 RA Ayadi A., Bordonne R., Canasses A., Madania A., Poch O.,  
 Tarassov I.A., Winsor B., Martin R.P.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 CC -! FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 SUBSTRATES.  
 CC -! CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 {RNA}(N).  
 CC -! SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.  
 CC THIS SUBUNIT IS THE SECOND LARGEST COMPONENT OF RNA POLYMERASE II.  
 CC -! SUBCELLULAR LOCATION: Nuclear.  
 CC -! MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
 FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 III FOR 5S AND TRNA GENES.  
 CC -! SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC  
 CC EMBL; M15893; AAA68096.1; -;  
 CC EMBL; U55020; AAC49637.1; -;  
 CC EMBL; Z75059; CAA99357.1; -;  
 CC PIR; A25884; A25884.  
 CC SGD; S0005677; RPB2.  
 CC InterPro; IPR001572; RNA\_pol\_B.  
 CC Pfam; PF00562; RNA\_pol\_B; 1.  
 CC PROSITE; PS01166; RNA\_POL\_BETA; 1.  
 CC Transferrase; DNA-directed RNA polymerase; Transcription; Zinc;  
 KW Zinc-finger; Nuclear protein.  
 FT ZN FING 1163 1185 C4-TYPE (POTENTIAL).  
 FT CONFLICT 1003 1006 AEGI -> RREY (IN REF. 1).  
 SQ SEQUENCE 1224 AA; 138751 MW; BABD03212C0A583E CRC64;  
 Alignment Scores:  
 Pred. No.: 2.76 Length: 1224  
 Score: 91.50 Matches: 49  
 Percent Similarity: 32.98% Conservative: 15  
 Best Local Similarity: 25.26% Mismatches: 63  
 Query Match: 6.18% Indels: 67  
 DB: 1 Gaps: 9  
 US-10-062-624-41 (1-840) x RPB2\_YEAST (1-1224)  
 QY 130 AAGTACATCCAGTATATACACTTTAGAAAAATCTCTGCTGAAGAACTCCTATTAT 189  
 Db 865 LysTyRGlyMetSerIleThr-----GluThrPheGluLysPro---Gln 878  
 QY 190 GGACACAAATCTCTCACTATAAAAAAGTTTCGGACTAAAGAAAGATGGTGATATAACAAA 249  
 Db 879 ArgThrAsnThrLeuArg-----MetLysHisGlyThrTyRAspLys 892  
 QY 250 AAAGACGATTTTACAAAGAGTAGCTCCAGGCATTGATTTTCAAAATAACTTAATATCAGA 309  
 Db 893 LeuAspAspGlyLeuIleAlaProGlyVal-----ArgValSerGly 907



Query Match:	6.22%	Indels:	65
DB:	1	Gaps:	11
US-10-062-624-41 (1-840) x NUSM_ASCSU (1-547)			
QY 732	TATCTTCTCAAATTTATTACCAATAAGCCATGGTAGTAGTATCCACCAATAAATGCAGAGAC	673	
Db 348	TyrIleLeuGluPhePhePheSerAenPhePheMetVal---ValPheAlaCysMetPhe	366	
QY 672	TTCTGGTGTCATAGGGTAACATAATACCTATTATTTCTCTGGTAAGCAAATTTTAGATTGAG	613	
Db 367	PhePheSerValPheLeuThrPheGlyTyr-----	376	
QY 612	GTCTTTAAAAATAGTATAGATCTGCTCCTATACCTGCACATGCATATGG-----	562	
Db 377	-----SerTyrArg-----LeuTyrPheGlyPhe	384	
QY 561	-----TACGAAGA-----TACTCCTTCAGCTCTAATGTATACCAAGTATTAAAC	517	
Db 385	PheMetSerPheSerArgProValPheCysPheSerSerValValMetAenPheLeu	404	
QY 516	CATCAATGACATAAAAGTTATGCGCTCATTTTAAAGTACTACATA-----TTGCTG	466	
Db 405	SerLeuLeuLeuValLeuPheSerIlePhePheIleTyrTrpMetAenPheAenMetLeu	424	
QY 465	ATCTTCCATTGCATCTTTACAGAGATAAATGCAAAATGTTTATAGTATTCACCAATTACGT	406	
Db 425	CysMetProCysLeuPheLeuTyr-ValAspPhePheValProLeuPhePheValValMet	444	
QY 405	ATCATTGTTATCGGTGTTTTTGGATTAATAATGTTTGATATGCAGCTTCAGTTCTATTCT	346	
Db 445	IleMetValValGlyPheLeuCysValLeuLeuLeuLeuGluPheValTyrLys---	463	
QY 345	TGGTCCGTCCATAGAGTAACCAATACTTCCTGAAATCCTGATATTAAAGTTATTTTGA	286	
Db 464	-----PheLeuValAspPhePhe	469	
QY 285	ATCAATGCCCTGGAGCTACTCTTTGAAATCGTCTTTTTTGTATATACCACTTTCTT	226	
Db 470	AlaLysGlyTyrValTyrGlyLeuLys-----AanTyrLysPhePheAspLeuPheLeu	487	
QY 225	TAGTCCGAAC-----TTTTTATGAGAGAAATTTGTTCCATTAAATAGG	181	
Db 488	GlyGlyIleAenSerLeuGlyValThrPhePheSer-----PheThr	501	
QY 180	AGTTTCTTCAGCAGAGAATTTCTAAAGTGTGATATACTCGATTGTACTTTGCACATA	121	
Db 502	GlyPheTyrSerAenSerTyrMetLysSerLeuTyrPheAenSerValValIleValLeu	521	
QY 120	GTAGAAGCCTCTTTTGTATC-----ATTAGTTCTTGAACCATCAGGATTCGCAA	70	
Db 522	ValLeuPhePhePheLeuValTyrGlyCysIleLeuSerLeuLysTyrAlaLeuCysLys	541	
QY 69	AGACTGATA	61	
Db 542	ArgMetIle	544	
RESULT 13			
VAC2_HELPY			
ID_VAC2_HELPY	STANDARD;	PRT;	1287 AA.
AC Q48245;			
DT 01-NOV-1997 (Rel. 35, Created)			
DT 01-NOV-1997 (Rel. 35, Last sequence update)			
DT 01-NOV-1997 (Rel. 35, Last annotation update)			
DE Vacuolating cytotoxin precursor.			
GN VACA.			
OS Helicobacter pylori (Campylobacter pylori).			
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;			
OC Helicobacter.			
OX NCBI_TaxID=210;			
RN [1].			
RC SEQUENCE FROM N.A.			
RP STRAIN=ATCC 49503 / 60190;			



QY	46	----	-----ATGCAATCTTACCATATCAGTCTTTTGAGAT	78
Db	549	AsnGluLeuIleValIyethrAsnGlyIleSerValGlyGluIyThrHisPheSerGlu	568	
QY	79	CTGTAGGTTCAAGTCAATATGATAAC	105	
Db	569	AspIleGlySerGlnSerArgIleasnThrValArgLeuGluThrGlyThrArgSerIle	588	
QY	106	-----AAAGAAGC-----TTCTACATT	123	
Db	589	PheSerGlyGlyValIyephelySerGlyGluLysLeuValIleAsnAspPheTyTr	608	
QY	124	AGT-----GCAAGTACAAATCCAAAGTATATACACACTTTAGAAAAATTC	165	
Db	609	SerProTrpAsnTyPheAspAlaArgAsnValLysAsnValGluIleThrArgLysPhe	628	
QY	166	TCGTCTGAAGAAACTCCATTAAAT-----GGAAACAATCTCTCACTAAAAAAGTTTTTC	219	
Db	629	---AlaSerSerThrProGluAsnProTrpGlyThrSerLysLeuMetPheAsnAsnLeu	647	
QY	220	GGACTAAGAAAGAGTGGTATATAACAAAAAAGACGATTTTACAGAGTAGTCCAGGC	279	
Db	648	ThrLeuGlyGlnAsnAla---ValMetAspTySerGlnPheSerAsnLeuThrIleGln	666	
QY	280	ATTGATTTTCAAAATAACTTAATATACAGGATTTTCAGAAAGTATTGGTTACTCTATGGAC	339	
Db	667	GlyAspPheIleAsnAsn-----GlnGlyThrIleAsnTyLeuValArg	681	
QY	340	GGACCAAGAATAGAACTT-----GAAGCTCATATCAACAATTTAATCCAAA	387	
Db	682	GlyGlyLysValAlaThrLeuAsnValGlyAsnAlaalaMetMetPhe-----	698	
QY	388	AACACCGATAACAATGATCTACTGATAATGGTGAA---TACTATAACATTTTGCATTA---	441	
Db	699	-----AsnAsnAspIleaspSerAlaThrGlyPheTyLysProLeuIleLysIle	715	
QY	442	----TCTCGTAAAGATGCAATGGAGATCAGCAATATGTACTTAAATAATGACGCATA	498	
Db	716	AsnSerAlaGlnAspLeuIleLysAsnThrGluHisValLeuLeuLysAlaLysIleIle	735	
QY	499	ACTTTTATGCTTCATTCATGCTTAATACTTGCTCATCACCATTCAGCTGAAGAGGATATCTTTC	558	
Db	736	GlyTyGlyAsnValSerThrGlyThr-----AsnGlyIleSerAsn	749	
QY	559	GTA	561	
Db	750	Val	750	

RESULT 15	
HLYA_VIBCH	
ID	HLYA_VIBCH STANDARD; PRT; 741 RA.
AC	P09545; Q99290; Q9KMD9; Q56641; Q08290; Q08197;
DC	01-MAR-1989 (Rel. 10, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Hemolysin precursor.
GN	HLYA OR VCA0219.
OS	Vibrio cholerae.
OC	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
NCBI_TaxID=666;	
[1]	SEQUENCE FROM N.A.
RP	STRAIN=E1 Tor O17 / Serotype O1;
RC	MEDLINE=89013889; PubMed=3050359;
RA	Alm R.A.; Stroehner U.H., Manning P.A.;
RT	"Extracellular proteins of Vibrio cholerae: nucleotide sequence of
RT	the structural gene (hlyA) for the haemolysin of the haemolytic E1
RT	Tor strain O17 and characterisation of the hlyA mutation in the non-
RT	haemolytic classical strain 569B.";
RL	Mol. Microbiol. 2:481-488(1988).
[2]	
RN	REVISONS.
RP	

RC STRAIN=El Tor O17 / Serotype O1;  
RA Manning P.A.;  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
[3]  
RN SEQUENCE FROM N.A.  
RP STRAIN=El Tor RV79;  
RX MEDLINE=88226910; PubMed=2453464;  
RA Rader A.E., Murphy J.R.;  
RT "Nucleotide sequences and comparison of the hemolysin determinants of  
RT Vibrio cholerae El Tor RV79 (Hly-) and RV79 (Hly+) and classical  
RT 569B (Hly-)";  
RL Infect. Immun. 56:1414-1419 (1988).  
[4]  
RN SEQUENCE FROM N.A., AND SEQUENCE OF 26-45 AND 158-177.  
RP STRAIN=El Tor N86 / Serotype O1;  
RX MEDLINE=91071898; PubMed=2174833;  
RA Yamamoto K., Ichinose Y., Shinagawa H., Makino K., Nakata A.,  
RA Iwanaga M., Honda T., Miwatani T.;  
RT "Two-step processing for activation of the cytotoxin/hemolysin of  
RT Vibrio cholerae O1 biotype El Tor: nucleotide sequence of the  
RT structural gene (hlyA) and characterization of the processed  
RT products";  
RL Infect. Immun. 58:4106-4116 (1990).  
[5]  
RN SEQUENCE FROM N.A.  
RP STRAIN=El Tor N16961 / Serotype O1;  
RX MEDLINE=20406833; PubMed=10952301;  
RA Heidelberg J.F., Eelsen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
RA Emdonald A.M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
RA McInaive L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
RA Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
RT cholerae";  
RL Nature 406:477-483 (2000).  
CC -1- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD  
CC CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY  
CC DEFINED.  
CC -1- SUBCELLULAR LOCATION: IN THE HEMOLYTIC BIOTYPE EL TOR THE 80 kDa  
CC HEMOLYSIN PRECURSOR IS SECRETED AND EXTRACELLULARLY CLEAVED,  
CC PRODUCING ACTIVE 65 KDA HEMOLYSIN.  
CC -1- INDUCTION: BY HEMOLYSIN B CYTOPLASMIC PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE AHH1/ASH4/HLYA/VVHA FAMILY OF  
CC HEMOLYSINS.  
CC -1- SIMILARITY: CONTAINS 1 RICIN B-TYPE LECTIN DOMAIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial/  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announcement/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; Y00557; CAA68637.1; --  
DR EMBL; M35855; AAA27528.1; --  
DR EMBL; X51746; CAA36035.1; --  
DR EMBL; AE004362; AAF96131.1; --  
DR TIGR; VCA0219; --  
DR InterPro; IPR001776; Aerolysin.  
DR InterPro; IPR001340; Hemlyan pore.  
DR InterPro; IPR000772; Ricin\_B\_lectin.  
DR Pfam; PF00652; Ricin\_B\_lectin; 2.  
DR Pfam; PF01117; Aerolysin; 1.  
DR SMART; SM00458; RICIN; 1.  
DR PROSITE; PS50231; RICIN B LECTIN; 1.  
KW Hemolysis; Toxin; Lectin; Signal; Complete proteome.  
FT SIGNAL 1  
FT PROPEP 26 157  
FT CHAIN 158 741 HEMOLYSIN.  
FT DOMAIN 484 575 RICIN B-TYPE LECTIN.  
FT FT

FT VARIANT 245 741 MISSING (IN THE NONHEMOLYTIC STRAIN 559B)

FT VARIANT 453 453 S -> F (IN STRAIN N16961)

FT CONFLICT 115 115 T -> S (IN REF. 3)

FT CONFLICT 562 562 G -> S (IN REF. 3)

SQ SEQUENCE 741 AA; 81961 MW; C99FAE2A01F37CCE CRC64;

Alignment Scores:

Pred. No.: 4.37 Length: 741

Score: 89.00 Matches: 68

Percent Similarity: 34.90% Conservative: 36

Best Local Similarity: 22.82% Mismatches: 90

Query Match: 6.01% Indels: 104

DB: 1 Gaps: 18

US-10-062-624-41 (1-840) x HLYA\_VIBCH (1-741)

QY 82 STAGGTTCAAGAACTAATAACAAA----- 108

DB 222 ValGlySerAlaThrProaspAlaLysIleValArgIleSerLeuAspAspSerThr 241

QY 109 ---GAAGGCTTCACATTAGT----- 108

DB 242 GlyAlaGlyIleHisLeuAsnAspGlnLeuGlyTyrArgGlnPheGlyAlaSerThr 261

QY 139 CCAAGTATATACACTTGAATAATCTCTGCTGGAAGAACTCCT----- 183

DB 262 ThrLeuAspAlaTyrPheArgGluTrpSerThrAspAlaIleAlaGlnAspTyrArgPhe 281

QY 184 ---ATTATGGAACAAATCTCTCACTAA-----AAAGTTTCGGACTAAAGAAAGAT 234

DB 282 ValPheAsnAlaSerAsnLysAlaGlnIleLeuLysThrPheProVal----- 298

QY 235 GGTGATATAACAAAAGACGATTTTACAGAGTAGTCCAGGCGATTGTTTCAAAT 294

DB 299 AspAsnIleAsnGluLys-----PheGluArg 307

QY 295 AACTTAATATCAGGATTT-----TCAGGAGTATTGGTTACTCTATGACGGA 342

DB 308 LysGluValSerGlyPheGluLeuGlyValThrGlyValGluValSerGlyAspGly 327

QY 343 CCAAGATAGAACTTCAAA-----GCTGCATATCAACA-----TTTAAT 381

DB 328 ProLysAlaLysLeuGluAlaArgAlaSerTyrThrGlnSerArgTrpLeuThrTyrAsn 347

QY 382 CCAAAAACACCGATAACATGATGATGATGATGATGATGATGATGATGATGATGAT 441

DB 348 ThrGlnAspTyrArgIleGluArgAsnAlaLysAsnAlaGlnAlaValSerPheThrTrp 367

QY 442 TCTCGTAAAGATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 492

DB 368 AsnArg-----GlnGlnTyrAlaThrAlaGluSerLeuLeuAsnArg 381

QY 493 GCGATAAATTTTATGTCATTGATGATGATGATGATGATGATGATGATGATGATGAT 552

DB 382 SerThrAsp-----AlaLeuTrpValAsnThr-----TyrProValAspValAsnArg 398

QY 553 TCTTCGTGACCATATGCA-----TGT 573

DB 399 SerProLeuSerTyrAlaSerPheValProLysMetAspValIleTyrLysAlaSerAla 418

QY 574 GCGAGTATAGGACAGATCTTATCACTATTTTAAAGACCTCAACTAAATTTGCTTAC 633

DB 419 ThrGluThrGlySerThrAspPheIleIleAspSerSerValIleArgProIleTyr 438

QY 634 CAAGGAAAATAGGTATTAGTTACCTATCACACAGAGAGTCTCTGCAATTTATGCTGA 693

DB 439 AsnGlyAlaTyrLysHisTyrTyrValValGlyAlaHisGlnSer----- 453

QY 694 TACTACCATGGGTATTGTTGTTAAATTTTGAAGAGATACCTGTTAACTCCTGTAGTA 753

DB 454 ---TyrHisGly-----PheGlu----- 458

QY 754 TTAATGATGCTCTCTCAA-----ACCACATCTGCTCTCAGTAACCTCTGAC 798

DB 459 -----AspThrProArgArgArgIleThrLysSerAlaSerPheThrValAsp 474

RESULT 16

AFAC\_ECOLI STANDARD; PRT; 859 AA.

AC P53517;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Outer membrane usher protein afac precursor.

GN AFAC.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI\_TaxID=562;

RN (1) SEQUENCE FROM N.A.

RC STRAIN=A30;

RA MEDLINE=95095929; PubMed=8002584;

RT "Nucleotide sequence of the afimbrial-adhesin-encoding afa-3 gene cluster and its translocation via flanking IS1 insertion sequences.";

RL J. Bacteriol. 176:7601-7613(1994).

CC - FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF AFA-III AFIMBRIAL

CC - ADHESIN SUBUNITS ACROSS THE OUTER MEMBRANE.

CC - SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane (by similarity).

CC - SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.

CC - CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-9 IS THE INITIATOR.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

CC EMBL; X76688; CAA54117.1; -

CC InterPro; IPR000015; Fimb\_usher.

DR Pfam; PF00577; Usher; 1

DR PROSITE; PS01151; FIMBRIAL\_USHER; 1

KW Outer membrane; Transmembrane; Fimbria; Transport; Signal; Plasmid.

FT SIGNAL 1 28 POTENTIAL.

FT CHAIN 29 859 OUTER MEMBRANE USHER PROTEIN AFAC.

SQ SEQUENCE 859 AA; 93805 MW; 1PF7PE3715E8FB93 CRC64;

Alignment Scores:

Pred. No.: 4.41 Length: 859

Score: 89.00 Matches: 57

Percent Similarity: 36.73% Conservative: 44

Best Local Similarity: 20.73% Mismatches: 94

Query Match: 6.01% Indels: 80

DB: 1 Gaps: 14

US-10-062-624-41 (1-840) x AFAC\_ECOLI (1-859)

QY 64 CAGTCTTTTTCAGATCTCTGAGTTCAGAACTAATGATGATAACAAGAGCTTCTACATT 123

DB 351 GlnValPheThrValProPheThrThrProAlaIleAlaLeuArgGluGlyTyr----- 368

QY 124 AGTCAAAAGTACAATCCAAAGTATATACACATTTTAAAAATTC-----TCTGCTGAAGAA 177

DB 369 ---LeuLysTyrAsnValThrAlaGlyGlnTyrArgSerAspAlaValGluHis 387

QY 178 ACTCTTATATGACAAATCTCTCACTAAAAAGTTTTCGGACTAAAGAAAGATGCT 237

DB 388 ThrSerLeuGlyGlnValThrAlaMetTyrGlyLeuProTrpGlyLeuThrValTyrGly 407

QY 238 GATATAACAAAAAAGACGATTTTACAGAGTAGTCTCCAGGCGATTGTTTCAAAATAAC 297





Db	424	----	-LeuGlyLeuThrThrSerLeuThrArgSerAspArgHisTyrArgGluLeuSer	441
QY	238	GATATAACAAAAAGACGATTTTTACA	-----AGACTGCTCCAGCATT	282
Db	442	Giualalle---	AspAspAspTyrThrAppProThrLysSerThrTyrAlaLeuGlyLeu	460
QY	283	GATTTCAAAATACTTAATATCAGGATTT	-----TCAGGAAGTATTGGTTACTCTATG	336
Db	461	AsnTrpSerAsnSerIleLeuGlyGlyPheAsnIleSerGlyTyrIleThrTyrSerTyr	480	
QY	337	CACGGACCA	-----AGATAGAATTGAACTGCTCATATCAA	372
Db	481	AspGlyAspAsnAspSerSerAsnLeuAsnIleAsnTrpAsnLysAlaPheLysHisAla	500	
QY	373	-----CAATTTAATCCAAAAACACCGATAACCAATCATACTGATTAAT	414	
Db	501	ThrValSerValAsnTrpGlnHisGlnLeuSerAlaSerGluAsnAsnGluAspAspGly	520	
QY	415	GGTGAATACTATTAACATTTTGGCATTA	441	
Db	521	AspLeuphetyrValAsnIleSerIle	529	
RESULT 20				
C131.DROME				
ID	C131 DROME	STANDARD;	PRT;	492 AA.
AC	Q9VFJ0;			
DT	16-OCT-2001	(Rel. 40, Created)		
DT	16-OCT-2001	(Rel. 40, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	Probable cytochrome P450 313a1 (EC 1.14.-.-)	(CYPCCCXIIIAI).		
OS	CYP313A1 OR CG3360.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;			
OC	Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;			
OC	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
NCBI	taxid=7227;			
RN	[1]	SEQUENCE FROM N.A.		
RC	STRAIN=Berkeley;			
RC	MEDLINE=20196006;	PubMed=10731132;		
RX	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.G., Wortman J.R., Yeung M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers Y.-H.C., Blaziej R.G., Champagne M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,			
RA	Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,			
RA	Baillet R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Beeoon K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,			
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,			
RA	Fowler C., Gabriellian A.E., Garg N.S., Galbart W.M., Glasner K.,			
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,			
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,			
RA	Jajali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,			
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,			
RA	Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,			
RA	Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,			
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,			
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,			
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,			
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,			
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,			
RA	Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,			
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,			
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,			
RA	Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,			







Db 425 ArgPheArgPheArgLeuThrValGlyGluProSerPheSerIleThrArgThrArgVal 444  
 QY 676 TCTGATTTATGGT-----GGTACTACCATGGC 705  
 Db 445 SerArgLeuTyrglyLeuProAlaAlaAsnProAsnAsnGlnArgGluTyrglyGlu--- 463  
 QY 706 GTTATTGGTAATAAATTTGAGAAGATACCTGTAATAACTCCGTAGTATTAAATGATGCT 765  
 Db 464 ---IleSerGlyArgPheSer-----LeuIleSerLeuValProSerAsnAspAsp 479  
 QY 766 COTCAAACC-----ACATCTGCTTCAGTAACCTTTGACGTTGGATGACTTTGCGCGGA 816  
 Db 480 TyrGlnThrProIleMetAsnSerValThrValArgGlnAspLeuGluArgGlnLeuGly 499  
 QY 817 GAAATGGAGTGGTTC 834  
 Db 500 GluLeuArgAspGluPhe 505

## RESULT 23

CCA\_BUCAI STANDARD; PRT; 414 AA.  
 AC P57169;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE tRNA nucleotidyltransferase (EC 2.7.7.25) (tRNA adenylyltransferase)  
 DE (tRNA CCA-pyrophosphorylase) (CCA-adding enzyme).  
 GN CCA OR BU061.  
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum  
 symbiotic bacterium).  
 OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.  
 OX NCBI\_TaxID=118099;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Tokyo 1998;  
 RX MEDLINE=20445173; PubMed=10993077;  
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;  
 RT "Genome sequence of the endocellular bacterial symbiont of aphids  
 Buchnera sp. APS.";  
 RL Nature 407:81-86(2000).  
 CC -!- FUNCTION: THIS ENZYME CARRIES OUT SYNTHESIS OF THE TRNA CCA  
 CC TERMINUS WITHOUT THE DIRECTION OF A TEMPLATE USING THE MULTIPLE  
 CC ACCEPTING AND DONATING SUBSITES WITHIN ITS ACTIVE SITE  
 CC (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: ATP + {tRNA}(N) = diphosphate + {tRNA}(N+1).  
 CC -!- SIMILARITY: BELONGS TO THE TRNA NUCLEOTIDYLTRANSFERASE / POLY(A)  
 CC POLYMERASE FAMILY.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

DR EMBL; AP001118; BAB12784.1; -  
 DR InterPro; IPR002646; PolyA\_pol.  
 DR Pfam; PF01743; PolyA\_pol; 1.  
 KW Transferase; Nucleotidyltransferase; ATP-binding; RNA-binding;  
 FT tRNA processing; Complete proteome.  
 FT NP BIND 68 73 ATP (POTENTIAL)  
 SQ SEQUENCE 414 AA; 48811 MW; 8CBF46515A47D4B7 CRC64;

## Alignment Scores:

Pred. No.:	6.89	Length:	414
Score:	86.50	Matches:	34
Percent Similarity:	45.59%	Conservative:	28
Best Local Similarity:	25.00%	Mismatches:	57
Query Match:	6.01%	Indels:	17
DB:	1	Gaps:	6

US-10-062-624-41 (1-840) x CCA\_BUCAI (1-414)

QY 555 AGTACTCCTTCAGCTGTAATGTCATAGCAGTATTAAACCATCAATGACATAAAAGTTAT 496  
 Db 251 ArgPheSerTyrglyLeuSerGlnPheLeu---SerValAsn---GlnIleAspArgAsnTyr 268  
 QY 495 GCC---GTCATTTTAAAGTACTACATATGCTGATCTTCCATTCGATCTTTTACGAGATAA 439  
 Db 269 SerIlePhePheAspSerTyrglyAlaAlaSerIleIleHisSerVal----- 284  
 QY 438 TGCAAAATGTTTATAGTATTCACCATATCAGTATCATGTTATCGGTGTTTTTGGATT 379  
 Db 285 CysIysArgPheIleProSerTyrglyLeuAspIleAlaVal----- 299  
 QY 378 AAATTTGTTGATGACGCTTCAGTCTTCACTTCTTGGTCCGTCCTAGTAGTAACCAATACT 319  
 Db 300 -----LeuAsnThrGlyPhePheLeuAsnThrIleHisTyrGlnSerSer--- 316  
 QY 318 TCCTGAAATCCTGATATTAAATTTTGAATAATCAATGCTGAGCTACTCTTTGTA 259  
 Db 317 -----LysAsnIleIleAsnLeuPheSerIysValAspAlaTdpArgLysProAspArg 334  
 QY 258 ATCGTCTTTTGTATATCACCATCTTCTTAGTCCGAAACTTTTGTAGTGAGAGA 199  
 Db 335 ValIysLysLeuAlaPheLeuSerAsnPheAsnPheLeuArgAsnPheLysSerGluPhe 354  
 QY 198 ATTTGTTCCATTAATAGGAGTTTCTTCAGCAGAGAAATTTCTAAAGTG 151  
 Db 355 PheCysIleLysSerGlyCysPheLeuGluLysCysPheSerValVal 370

RESULT 24  
 CYSJ\_BUCAI STANDARD; PRT; 601 AA.  
 ID CYSJ\_BUCAI  
 AC P57503;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Sulfite reductase [NADPH] flavoprotein alpha-component (EC 1.8.1.2)  
 DE (SR-FP)  
 GN CYSJ OR BU428.  
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum  
 symbiotic bacterium).  
 OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.  
 OX NCBI\_TaxID=118099;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Tokyo 1998; PubMed=10993077;  
 RX MEDLINE=20445173; PubMed=10993077;  
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;  
 RT "Genome sequence of the endocellular bacterial symbiont of aphids  
 Buchnera sp. APS.";  
 RL Nature 407:81-86(2000).  
 CC -!- FUNCTION: THIS ENZYME CATALYZES THE 6-ELECTRON REDUCTION OF  
 CC SULFITE TO SULFIDE. THIS IS ONE OF SEVERAL ACTIVITIES REQUIRED  
 CC FOR THE BIOSYNTHESIS OF L-CYSTEINE FROM SULFATE. THE FLAVO-  
 CC PROTEIN COMPONENT CATALYZES THE ELECTRON FLOW FROM NADPH ->  
 CC FAD -> FMN TO THE HEMOPROTEIN COMPONENT (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: H(2)S + 3 NADP(+) + 3 H(2)O = sulfite + 3  
 CC NADPH.  
 CC -!- COFACTOR: THIS SUBUNIT IS A FLAVOPROTEIN THAT BINDS ONE FMN AND  
 CC ONE FAD PER CHAIN (BY SIMILARITY).  
 CC -!- SUBUNIT: ALPHA(8)-BETA(4). THE ALPHA COMPONENT IS A FLAVOPROTEIN,  
 CC THE BETA COMPONENT IS A HEMOPROTEIN (BY SIMILARITY).  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL; AP001119; BAB13126.1; -  
 DR HSP; P00388; IAMO.



```

RX MEDLINE=92154671; PubMed=1339306;
RA Peterson C.L., Herskowitz I.;
RT "Characterization of the yeast SWI1, SWI2, and SWI3 genes, which
   encode a global activator of transcription.";
RL Cell 68:573-583(1992).
CC -!- FUNCTION: INVOLVED IN TRANSCRIPTIONAL ACTIVATION. THE SWI/SNF
CC COMPLEX IS REQUIRED FOR THE INDUCED EXPRESSION OF A LARGE NUMBER
CC OF GENES. THIS COMPLEX ALTERS CHROMATIN STRUCTURE TO FACILITATE
CC BINDING OF GENE-SPECIFIC DEDICATED TRANSCRIPTION FACTORS.
CC -!- SUBUNIT: COMPONENT OF THE SWI/SNF GLOBAL TRANSCRIPTION ACTIVATOR
CC COMPLEX.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 ARID DOMAIN.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X12493; CAA31013.1; -.
CC EMBL; U33335; AAB68089.1; -.
CC PIR; S05728; TNBYR6.
CC TRANSFAC; T01279; -.
CC SGD; S0005937; SWIL.
CC InterPro; IPR001606; ARID.
CC Pfam; PF01388; ARID; 1.
CC SMART; SM00501; BRIGHT; 1.
CC Transcription regulation; Activator; DNA-binding; Nuclear protein;
CC Zinc-finger.
CC DOMAIN 5 65 ASN/THR-RICH.
CC FT DOMAIN 337 385 GLN-TYPE.
CC FT ZN FING 1241 1258 C4-TYPE.
CC SQ SEQUENCE 1314 AA; 147938 MW; F442D5A82013CDBD CRC64;

Alignment Scores:
Pred. No.: 8.17 Length: 1314
Score: 86.00 Matches: 35
Percent Similarity: 40.2% Conservatives: 23
Best Local Similarity: 24.3% Mismatches: 54
Query Match: 5.81% Indels: 32
DB: 1 Gaps: 7

US-10-062-624-41 (1-840) x SWIL_YEAST (1-1314)
OY 16 ATTCCTAGTAAGAGCGGTTAATCTCAATATGTAATCTTACCATATCAGTCTTTTGA 75
DB 164 lleuAlaLysAsnSerIleleasp---SerSerAsnLeuProleuGlnAlaGln 182
OY 76 GATCTGTAGGTTCAAGAACTAATGATACAAAGAGGCTTC-----TACATT 123
DB 183 GlnLeuTyGlyGlyAsnAsnAsnAsnSerThrGlyIleAlaAsnAspValle 202
OY 124 ACTGCAAGTACAATCCAGTATATACACTTTAGAAATTCCTGCTCAAGAACTCT 183
DB 203 ThrProHisPheIleThrAsnValGlnSerIleSerGlnAsnSerSerSerThr 222
OY 184 ATTAATGGAACAAATCTCTC-----ACTAAAGAGTTTTC-----GGA 222
DB 223 AsnThrAsnSerAsnSerThrProAsnAlaAsnGlnPheLeuProPheAsnAsn 242
OY 223 CTAAGAAGATGGTGATATACAAAGAGAGATTTTACAAGAGTAGCTCCAGGCATT 282
DB 243 AlaSerAsnAsnGlyAsnLeuThr-----
OY 283 GATTTCCTCAAAATTAATATCAGGATTTTTCAGGAGTATGTTGTTACTCTATGAC 342
DB 251 -----SerAsnGlnLeulleSerAsnTyAlaAlaSer-----AsnSerMet 265
OY 343 CCAAGATAGACTTGAAGCTGATATCAACATTTTAATCCAAAAACACCCGATACAT 402
DB: 1 Gaps: 1
```

```

Db 266 -----ArgSerSerSerAlaSerAsnGluPheValProAsnThrSerAspAsnAsn 282
OY 403 GATACTGATAAT 414
DB 283 AsnAsnSerAsn 286

RESULT 26
YCCE_ECOLI
ID YCCE_ECOLI STANDARD; PRT; 418 AA.
AC P36661; P75889;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ycce (ORF-D).
GN YCCE OR B1001.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1] TaxID=562;
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Iano M., Horuchi T.;
RA "A 718-kb DNA sequence of the Escherichia coli K-12 genome
   corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [3]
RP SEQUENCE OF 1-242 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=94134696; PubMed=8302830;
RA Ueguchi C., Kakeda M., Yamada H., Mizuno T.;
RT "An analogue of the DnaJ molecular chaperone in Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:1054-1058(1994).
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000202; AAC74086.1; -.
CC EMBL; D90736; BAA36143.1; -.
CC EMBL; D90737; BAA35768.1; -.
CC EMBL; D16500; -. NOT_ANNOTATED_CDS.
CC Ecogene; EG12196; ycce.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 418 AA; 48355 MW; D1D52AD49708F8D9 CRC64;

Alignment Scores:
Pred. No.: 8.39 Length: 418
Score: 85.50 Matches: 40
Percent Similarity: 42.76% Conservatives: 25
Best Local Similarity: 26.32% Mismatches: 56
Query Match: 5.77% Indels: 31
DB: 1 Gaps: 9
```





RL Nucleic Acids Res. 24:4420-4449(1996).  
 CC -!- SIMILARITY: BELONGS TO THE MG067 / MG068 / MG395 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE000025; AAB95904.1; -  
 DR InterPro: IPR002414; DUF30/31.  
 DR Pfam: PF01727; DUF30; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 347 AA; 40215 MW; 2C457C1401861B88 CRC64;  
 -----  
 Alignment Scores:  
 Pred. No.: 11.1 Length: 347  
 Score: 84.00 Matches: 33  
 Percent Similarity: 34.92% Conservative: 11  
 Best Local Similarity: 26.19% Mismatches: 38  
 Query Match: 5.67% Indels: 44  
 DB: 1 Gaps: 5  
 -----  
 US-10-062-624-41 (1-840) x YF86\_MYCPN (1-347)  
 QY 238 GATATAACAAAAAGACGATTTTACAGAGTAGCTCCAGCGATTGATTTCAA----- 291  
 DB |||:|||||  
 QY 221 AspleuSerLysLysSerGlyLeuTyHisLysPheLeuGlyLysTyGlnLeuTyR 240  
 DB -----  
 QY 292 -----AATACTTAATATCAGGATTTTCAGGAAGT----- 321  
 DB |||:|||||  
 QY 241 GlyLysSerIleValIleSerAspLeuAsnLeuSerSerGlySerGlySerLeu 260  
 QY 322 -----ATTGGTACTCTATGAGCGACCGACCAAGATAGAA 354  
 DB |||:|||||  
 QY 261 LeuAsnAspArgLysGlnIleValGlyLeuTyPheGlyValAspGlyProLysAspGlu 280  
 QY 355 CTTGAAGCTGATATCAACATTTTAATCCAAAAACACCGATAACATGATCTGATAAT 414  
 DB |||:|||||  
 QY 281 Leu-----GlyPheSerGlnLeuLeuArgTrpGlnAlaLysAsnAspGluGluLys 298  
 QY 415 GGT-----GAATCTATAAATACAT 432  
 DB |||:|||||  
 QY 299 AspSerValAlaTyAspLeuIlePheGlyAsnLysAsnThrThrLysTyTyrAlaGln 318  
 QY 433 TTTGCATATCTCTGTA-----GATGCAATGGAAGATCAGCAA 471  
 DB |||:|||||  
 QY 319 PheAlaLysGluHisLysThrHisLeuTyGluGlnIleAspArgSerAsnAspGlnGln 338  
 QY 472 TATGTAGTACTTAAATAAT 489  
 DB |||:|||||  
 QY 339 PheThrPheValLysAsn 344  
 DB |||:|||||  
 RESULT 30  
 ID OPAP\_NEIGO STANDARD; PRT; 234 AA.  
 AC Q04879;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Opacity protein OPA56 precursor (Fragment).  
 GN OPAP.  
 OS Neisseria gonorrhoeae.  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=485;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MS11 / F3;  
 RX MEDLINE=93178439; PubMed=8440254;  
 RA Kupsch E.-M., Knepper B., Kuroki T., Heuer I., Meyer T.F.;  
 RT "Variable opacity (Opa) outer membrane proteins account for the cell

tropisms displayed by Neisseria gonorrhoeae for human leukocytes and epithelial cells.";  
 RT EMO J. 12:641-650(1993).  
 CC -!- FUNCTION: IMPLICATED IN A NUMBER OF ADHERENCE FUNCTIONS. OPA  
 CC PROTEINS ARE IMPLICATED IN PATHOGENESIS AND ARE SUBJECT TO PHASE  
 CC VARIATION.  
 CC -!- SUBCELLULAR LOCATION: Outer membrane.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z18934; CAA79367.1; -  
 DR PIR: S28620; S28620.  
 DR InterPro: IPR003394; Porin\_opacity.  
 DR Pfam: PF02462; Opacity; 1.  
 KW Outer membrane; Multigene family; Signal.  
 FT NON TER 1 1  
 FT SIGNAL <1 1 POTENTIAL.  
 FT CHAIN 234 234 OPACITY PROTEIN OPA56.  
 FT NON TER 234 234  
 SQ SEQUENCE 234 AA; 26868 MW; 5175C6606839EFB8 CRC64;  
 -----  
 Alignment Scores:  
 Pred. No.: 12 Length: 234  
 Score: 83.50 Matches: 63  
 Percent Similarity: 31.10% Conservative: 30  
 Best Local Similarity: 21.07% Mismatches: 86  
 Query Match: 5.64% Indels: 120  
 DB: 1 Gaps: 14  
 -----  
 US-10-062-624-41 (1-840) x OPAP\_NEIGO (1-234)  
 QY 100 GATAACAAAGAGCTCTTACATTTAGTCAAGTACAATCAAGTATATCACACTTTAGA 159  
 DB |||:|||||  
 QY 4 AsphHisGlyArgGlyProTyValGlnAlaAspLeuAlaTyAlaTyGluHisIleThr 23  
 QY 160 AATTCTCTGCTGAGAACTCTTATTAAATGGAACAAATCTCTCACATAAAAAAGTTTTC 219  
 DB |||:|||||  
 QY 24 HisAspTyProGluGlnThr----- 30  
 QY 220 GGACTAAGAAAGATGGTGATATATAACAAAAACAGCATTTTCAACAGAGTAGCT----- 273  
 DB |||:|||||  
 QY 31 GlyThrLysLysAspLysIleSerThrValSerAspTyPheArgAsnValArgThrHis 50  
 QY 274 -----CCAGCGATTTGATTTTCAAAATAACTTAATATCAGGATTTTTCAGGAAGTATT 324  
 DB |||:|||||  
 QY 51 SerIleHisProArgVal-----SerVal 58  
 QY 325 GGTACTCTATGGACGACCAAGATAGACTTGA-----GCTGCATATCAACAATTAAT 381  
 DB |||:|||||  
 QY 59 GlyTyAspPheGlyGlyTrpArgIleAlaAspTyAlaAspTyArgLysTrpAsn 78  
 QY 382 CCA-----AAAAACCGATAACAATCAT 405  
 DB |||:|||||  
 QY 79 AspAspLysTySerValAspIleLysGluLeuGluAsnLysAsnGlnAsnLysArgAsp 98  
 QY 406 -----ACTGATATGTGTAATCTACTATAAAACATTTTGCATTTATCTCGTAAAGATGCAATG 459  
 DB |||:|||||  
 QY 99 LeuLysThrGluAsnGlnGluAsnGlyThrPheHisAlaValSer----- 113  
 QY 460 GAAGATCAGCAATATGATAGTACTTTAAAAATACGCGCATAACTTTTATGTCATTGATGTT 519  
 DB |||:|||||  
 QY 114 -----SerLeuGlyLeu 117  
 QY 520 AATATTGCTATGACATTACAGCTGAAGAGTATCTTTTGTACCATATGATGTCAGGT 579  
 DB |||:|||||  
 QY 118 SerAlaValTyAspPheLysLeuAsnAsp-----LysPheLysProTyR----- 132

QY 580 ATAGGAGAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGA 639  
 Db |||||  
 133 IleGlyAla---ArgValalaTy---Gly 140

QY 640 AAAATAGGTATTAGTTACCTTATCACACGAGAGTCTCTGCATTTATTTGGTGATACAC 699  
 Db |||||  
 141 HisValArgHisSerIleAspSerThrLysLysThrLysPheLeuThrSerSerTy 160

QY 700 CATGCGTATTGGTAATAATTGAGAAG----- 729  
 Db |||||  
 161 GlyGlyLeuAsnProThrValTyThrGluGluAsnThrGlnAsnAlaHisGlnSer 180

QY 730 -----ATACCTGTAATACTCTCTAGTATTAAATCATGCTCTCT 768  
 Db |||||  
 181 AsnSerIleArgValGlyLeuGlyValIleAlaGlyValGlyPheAspIleThrPro 200

QY 769 CAACACCATCTGCTTCAGTAACCTCTGACGTTGGATAC----- 807  
 Db |||||  
 201 Lys-----LeuThrLeuAspThrGlyTyArgTyHisTyTrpGlyArg 215

QY 808 -----TTGCGCGAGAAATTGGAATGAGTTTCACCTTC 840  
 Db |||||  
 216 LeuGluAsnThrArgPheLysThrHisGluAlaSerLeuGlyValArgTyArgPhe 234

RESULT 31  
 PEPT\_CLOPE  
 ID PEPT\_CLOPE STANDARD; PRT; 406 AA.  
 AC Q8XP58;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Peptidase T (EC 3.4.11.14) (Tripeptide aminopeptidase)  
 DE (Aminotripeptidase) (Tripeptidase).  
 GN PEPT OR CPE0025.  
 OS Clostridium perfringens.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1502;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=13 / Type A;  
 RX PubMed=11792842;  
 RA Shimizu T., Ohtani K., Hirakawa H., Oshima K., Yamashita A.,  
 RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;  
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic  
 flesh-eater.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
 CC -!- FUNCTION: Release the N-terminal amino acid from tripeptides. Has  
 a preference for tripeptides containing a N-terminal methionine or  
 leucine (By similarity).  
 CC -!- CATALYTIC ACTIVITY: Release of a N-terminal residue from a  
 tripeptide.  
 CC -!- COFACTOR: Zinc (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20A.

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AP003185; BAB79731.1; -  
 DR InterPro; IPR001261; ARGE\_DAPE\_CPG2.  
 DR InterPro; IPR002933; Peptidase\_M20.  
 DR Pfam; PF01546; Peptidase\_M20; 1.  
 DR PROSITE; PS00758; ARGE\_DAPE\_CPG2\_1; 1.  
 DR PROSITE; PS00759; ARGE\_DAPE\_CPG2\_2; 1.  
 KW Hydrolase; Aminopeptidase; Metalloprotease; Zinc; Complete proteome.  
 FT METAL 78 78 ZINC (BY SIMILARITY).  
 FT METAL 139 139 ZINC (BY SIMILARITY).

FT METAL 196 196 ZINC (BY SIMILARITY).  
 SQ SEQUENCE 406 AA; 45224 MW; A3E17B2A7F2FAEB7 CRC64;

Alignment Scores:  
 Pred. No.: 13.7 Length: 406  
 Score: 83.00 Matches: 46  
 Percent Similarity: 39.06% Conservative: 29  
 Best Local Similarity: 23.96% Mismatches: 54  
 Query Match: 5.60% Indels: 53  
 DB: 1 Gaps: 12

US-10-062-624-41 (1-840) x PEPT\_CLOPE (1-406)

QY 4 AATATAGAAATCTAGTAAGAGCGGTTATCTCATTAAATGTCAATCTTACCATAT 63  
 Db |||||  
 65 AsnCysAsnLysAspIleProLysIleGlyPheIleSerHisMetAspThrAlaPro--- 83

QY 64 CAGTCTTTGAGATCCTGTAGTTCAAGAACTAAT-----CAT 102  
 Db |||||  
 84 -----AspMetSerGlyLysAsnValAsnProLysIleValGluAsnTy 99

QY 103 AACAAAGAA-----GGCTTACATTAGTGCAGAAAGTACAATCCAGTATA 147  
 Db |||||  
 100 GlyLysAspIleGluLeuGlyAsnGlyThrLeuSerProSerPheSerProGluLeu 119

QY 148 TCACACTTTAGA-----AAATTCCTCTGAGAAACTCTTATTAATGGAACAAT 198  
 Db |||||  
 120 ProMetTyLysGlyGlnThrLeuIleThrAspGlyThrThrLeuLeuGlyAlaAsp 139

QY 199 -----TCTCTCACTAAAGAAAGTTTCGAGCTAAAG----- 228  
 Db |||||  
 140 AspLysAlaGlyIleAlaGluIleValThrAlaIleGluTyLeuIleAsnHisProGlu 159

QY 229 ---AAAGATGGTGATATAACAAAGAAAGACGATTTTACA-----AGAGTAGCTCCA 276  
 Db |||||  
 160 IleLysHisGlyAspIle-----LysIleGlyPheThrProAspGluGluIleGlyGlu 177

QY 277 GGCATTGATTTTCAAATAACTTAATATCAGATTTTCAGAGAGTATTGGTTACTCTATG 336  
 Db |||||  
 178 GlyAlaAspHisPheAsp-----ValGluGlyPheGlyAlaAspPheAlaTyThrLeu 195

QY 337 GACGACCAAGAATA---GAATCTTGAAGCTGCATATCAACAATTTAAATCCAAAAAACACC 393  
 Db |||||  
 196 AspGlyGlyArgIleGlyGluLeuGlu-----TyrGluAsnPheAsnAlaSerAla 213

QY 394 GATAACAATGATCTG-----ATAATGCTGAA 420  
 Db |||||  
 214 Lys-ValGluIleIleGlyLysAsnValHisProGlySerAlaLysGlyLysMetIleAs 233

QY 421 TACTATAACATTTTGCATTATCTCGTAAAGATG 454  
 Db |||||  
 233 nSerIleLeuValAlaHisGluPheValSerMet 244

RESULT 32  
 MURE\_STAM  
 ID MURE\_STAM STANDARD; PRT; 494 AA.  
 AC Q99V74;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE UDP-N-acetylmuramoylalanine-D-glutamate--2,6-diaminopimelate ligase  
 DE (EC 6.3.2.13) (UDP-N-acetylmuramyl-tripeptide synthetase) (Meso-  
 diaminopimelate-adding enzyme) (UDP-MurNAC-tripeptide synthetase).  
 GN MURE OR SAV1018 OR SA0876.  
 OS Staphylococcus aureus (strain N315).  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=158878; 158879;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Mure50 / ATCC 700699, and N315;  
 RX MEDLINE=21311952; PubMed=11418146;  
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,



QY	490	GACGCGATAACTTTTATGTCATTGATCGGTAAATACCTTGTCTATGACATTACGAGTGAAGGA	549
Db	241	AspAspSerPheSerGluTyrLeuArgThrValThrProTyrGluValPheSerTyrGly	260
QY	550	GTA-----TCITTCGTACCATATGCATGTGCAGGTATAGGA	595
Db	261	IleAspGluGluLaGlnPheMetAlaGlyAsnIleGlnGluSerLeuGlnGlyValSer	280
QY	586	CGACAGTCTTATCACTATTATTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAATA	645
Db	281	PheAspPheValThrProPheGlyThrTyrProValIysSerProTyrValGlyIysPhe	300
QY	646	GGTATTAGT 654	
Db	301	AsnIleSer 303	
RESULT 33			
FTSI_ECOLI			
ID	FTSI_ECOLI	STANDARD;	PRT; 588 AA.
AC	P04286;		
DT	20-MAR-1987 (Rel. 04, Created)		
DT	20-MAR-1987 (Rel. 04, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Peptidoglycan synthetase ftsI precursor (penicillin-binding protein 3) (PBP-3).		
DE	FTSI OR PBPP OR B0084 OR Z0094 OR EGS0088.		
GN	Escherichia coli, and		
OS	Escherichia coli O157:H7.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Escherichia.		
OX	NCBI_TaxID=562, 83334;		
[1]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=K12;		
RX	MEDLINE=83296957; PubMed=6350821;		
RA	Nakamura M., Maruyama I.N., Soma M., Kato J., Suzuki H., Horota Y.;		
RA	"On the process of cellular division in Escherichia coli: nucleotide		
RT	sequence of the gene for penicillin-binding protein 3.,"		
RT	Mol. Gen. Genet. 191:1-9(1983).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=K12;		
RX	MEDLINE=92334977; PubMed=1630901;		
RA	Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,		
RA	Isono K., Mizobuchi K., Nakata A.;		
RA	"Systematic sequencing of the Escherichia coli genome: analysis of		
RT	the 0-2.4 min region.,"		
RL	Nucleic Acids Res. 20:3305-3308(1992).		
[3]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=K12 / MG1655;		
RX	MEDLINE=97426617; PubMed=9278503;		
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,		
RA	Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,		
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,		
RA	Mau B., Shao Y.;		
RT	"The complete genome sequence of Escherichia coli K-12.,"		
RL	Science 277:1453-1474(1997).		
[4]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=O157:H7 / EDL933 / ATCC 700927;		
RX	MEDLINE=21074935; PubMed=11206551;		
RA	Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,		
RA	Rose D.J., Mayhew G.P., Evans P.S., Gregor J., Kirkpatrick H.A.,		
RA	Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,		
RA	Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,		
RA	Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,		
RA	Welch R.A., Blattner F.R.;		
RT	"Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.,"		
RL	Nature 409:529-533(2001).		
[5]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=O157:H7 / RIMD 0509952;		
RX			



```

Qy 343 CCAAGATAGAACTTGAAGCT---GCATATCAACAATTTAATCAAAACACCGATAAC 399
Db 239 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Qy 400 AAT-----CATACTGATATGCTGATACATAT----- 426
Db 256 LysalaGlusSerGlySerAlaValLeuValAsnThrGlyGluValLeuAlaMet 275
Qy 427 -----AAACATTTTGCATTTATCTCTGTAAGATGCAATGAA 462
Db 276 AlaAsnSerProSerTyrAsnProAsnAsnLeuSerGlyThrProLysGluAlaMetArg 295
Qy 463 GATCAGCAATATGATGACTTAAATAAGACGGCATACTTTTATGTCATGATGTTAAT 522
Db 296 AsnArgThrIleThrAspValPheGluProGlySerThrValLysProMetValValMet 315
Qy 523 ACTTGCTATGAC-----ATTACAGCTGAAGG-----GTATCTTCGTACCATATGCA 570
Db 316 ThrAlaLeuGlnArgGlyValValArgGluAsnSerValLeuAsnThrIleProTyrArg 335
Qy 571 TGTGAGGTATGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCT 630
Db 336 IleAsnGlyHisGluIleLysAspValAlaArgTyrSerGluLeuThrLeuThrGlyVal 355
Qy 631 TACCAA-----GGAAAAATAGGTATTAGT---TACCCATATCAACAGAGTCTCTGCA 681
Db 356 LeuGlnLysSerSerAsnValGlyValSerLysLeuAlaLeuAlaMetProSerSerAla 375
Qy 682 TTTATTGCTGATACCAT-----GGCGTTATTGGT 714
Db 376 LeuValAspThrTyrSerArgPheGlyLeuGlyLysAlaThrAsnLeuGlyLeuValGly 395
Qy 715 AATAAAATTGAGAAGATACCT----- 735
Db 396 GluArgSerGlyLeuTyrProGlnLysGlnArgTrpSerAspIleGluArgAlaThrPhe 415
Qy 736 -----GTAATAACTCTCTAGTATTAAATGATGCTCTCAAAACCACA 777
Db 416 SerPheGlyTyrGlyLeuMetValThrProLeuGlnLeu----- 428
Qy 778 TCTGCTTCAGTAACCTCTGAGTTGGATCTTTGGC 813
Db 429 ---AlaArgValTyrAlaThrIleGlySerTyrGly 439

RESULT 34
ID_ VP4_ROTFF6 STANDARD; PRT; 776 AA.
AC Q07416;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer capsid protein VP4 (Hemagglutinin) (Outer layer protein VP4)
DE [Contains: Outer capsid proteins VP5 and VP8].
DE S4.
GN Feline rotavirus (strain FRV-64).
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=39010;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94112374; PubMed=8284514;
RA Isegawa Y., Mochizuki M., Nakagomi T., Ueda S., Nakagomi O.;
RT "The VP4 gene sequence of a haemagglutinating strain of feline
RL rotavirus.";
RL Res. Virol. 144:371-374(1993).
CC -1- SUBCELLULAR LOCATION: Outer capsid.
CC -1- PTM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
CC OTHER PRODUCT IS VP5.
CC -1- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D14723; BAA03543.1; --
DR InterPro; IPR000416; Cap_VP4.
DR Pfam; PF00426; VP4; 1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 776
FT CHAIN 1 241
FT CHAIN 248 776
FT CARBOHYD 32 32
FT CARBOHYD 56 56
FT CARBOHYD 97 97
FT CARBOHYD 111 111
FT CARBOHYD 149 149
FT CARBOHYD 183 183
FT CARBOHYD 198 198
SQ SEQUENCE 776 AA; 86462 MW; 5B5A2A0CC4389ED1 CRC64;

Alignment Scores:
Pred. No.: 14.3 Length: 776
Score: 83.00 Matches: 69
Percent Similarity: 34.55% Conservative: 54
Best Local Similarity: 19.38% Mismatches: 127
Query Match: 5.60% Indels: 106
DB: 1 Gaps: 20

US-10-062-624-41 (1-840) x VP4_ROTFF6 (1-776)
Qy 43 TTAATGTCAATCTTACCATATCAGTCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGAT 102
Db 164 LeuTyrAlaValMetLysTyrGlyGlyArgIleHisThrTyrSerGlyGlnThrProAsn 183
Qy 103 AACAAAGAGCTTCTACATTAGTGCAAGTACATCCA---AGTATATCACACTTTAGA 159
Db 184 AlaThrThrGlyTyrTyrSerAlaThrAsnTyrAspSerValAsnMetThrThrPheCys 203
Qy 160 AAATTTC-----TCTGCTGAAGAAACT----- 180
Db 204 AspPheTyrIleIleProArgSerGluGluSerLysCysThrGluTyrIleAsnAsnGly 223
Qy 181 -----CCTATTATGAACAAATTTCTCTC-----ACTAAAAAGTT 216
Db 224 LeuProIleGlnAsnThrArgAsnIleValProLeuAlaLeuSerAlaArgAsnVal 243
Qy 217 TTCGACTAAAGAAGATGGTGAT-----ATAACAAA 249
Db 244 IleSerLeuLysAlaGlnSerAsnGluAspIleValValSerLysThrSerLeuTrpLys 263
Qy 250 AAAGACGATTTTACAAGA---GTAGCTCCAGGCATTGATTTTCAAAATAACTTAATATCA 306
Db 264 GluMetGlnTyrAsnArgAspIleThrIleArgPheLysPheAlaAsnSerIleValLys 283
Qy 307 GGATTTTTCAGGAAGTATTGGTTACTCT----- 333
Db 284 -----SerGlyGlyLeuGlyTyrLysTrpSerGluIleSerPheLysProAlaAsnTyr 301
Qy 334 -----ATGGACGGACCAAGATAAGTCTGAAGCTGATATCAACAATTT 378
Db 302 GlnTyrThrTyrMetArgAspGlyGluGluVal-----ThrAlaHisThrThrCys 318
Qy 379 AATCCAAAAACACCGATACATGATGATGATAATGCTGAATACTATAACATTTTGA 438
Db 319 SerValAsnGlyMetAsnAspPheSerPheAsnGlyGlySerLeuProThrAspPheVal 338
Qy 439 TTATCTCGTAAAGATGCAATGGAAGATGACAGCAATATAGTACTATAA-----AATGAC 492
Db 339 IleSerArgTyrGluValIleLysGluAsnSerTyrValTyrIleAspTyrTrpAspAsp 358
Qy 493 GGCATAACTTTT-----ATGTCATTGATGGTTAAT----- 522

```



```

Pred. No.: 14.4 Length: 878
Score: 83.00 Matches: 40
Percent Similarity: 36.81% Conservative: 27
Best Local Similarity: 21.98% Mismatches: 73
Query Match: 5.60% Indels: 42
DB: 1 Gaps: 6

US-10-062-624-41 (1-840) x OSB2_HUMAN (1-878).
QY 292 AATACTTAATATCAGGATTTTCAGGAAGTATGTGTTACTTATCGACGACCAAGAATA 351
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
373 AsnSerLeuGluArgAlaPheHisSerAlaProGlyArgProAlaAsnProSerLysSer 392
QY 352 GAACCTTGAGCTGCATATCAACAATTTTAAATCAAAAACACCGATAACAATGATGAT 411
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
393 PheIleGluGlySerLeu-----LeuThrProLysGlyGluAspSerGluGluAspGlu 410
QY 412 AATGCTGAATATATAAACAATTTTGCATTAATCTCTGTAAGATGCAATGAGATCAGCAA 471
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
411 AspThrGluTyrPhe-----AspAlaMetGluAspSerThr 422
QY 472 TATGTAGTACTT-----AAAAATGACGGCATAACT 501
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
423 SerPheIleThrValIleThrGluAlaLysGluAspSerArgLysAlaGluGlySerThr 442
QY 502 TTTATGTCATGATGGTTAATCTGTCATGACATACAGCTGAGGAGTATCTTCTGTA 561
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
443 GlyThrSerSerValAspTrpSerSerAlaAspAsnValLeuAspGlyAlaSerLeuVal 462
QY 562 CCATATGATGTCAGGATATAGGACGACATCTTATCACTATTTTAAAGACCTCAATCTA 621
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
463 Pro----- 463
QY 622 AAATTTGCTTACCAAGGAAATAGGTATTAGTTACCTTATCACACCAAGAGTCTTGCA 681
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
464 LysGlySerSerLysValLysArgValArgLysProAsnLysProAsnTyrSerLeu 483
QY 682 TTTATGCTGATACATCATGGCGTTTATGTTAATAAATTTGAGAAGATCACTGTATA 741
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
484 AsnLeuTrpSerIleMetLysAsnCysIleGlyArgGluLeuSerArgIleProMet--- 502
QY 742 ACTCTGTAGTATTAATGATGCTCTCAACACCATCTGCTTCAGTAACCTCTTGACGTT 801
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
503 ---ProValAsnPheAsnGlu---ProLeuSerMetLeuGlnArgLeuThrGluAspLeu 520
QY 802 GATAC 807
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
521 Glutyr 522

RESULT 36
OPAE_NEIGO
ID OPAE_NEIGO STANDARD; PRT; 234 AA.
AC Q04878;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE Opacity protein OPA55 precursor (Fragment).
GN OPAE.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_taxid=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MS11 / F3;
RX MEDLINE=93178439; PubMed=8440254;
RA Kupsch E.-M., Knepper B., Kuroki T., Heuer I., Meyer T.F.;
RT "Variable opacity (Opa) outer membrane proteins account for the cell
RT tropisms displayed by Neisseria gonorrhoeae for human leukocytes and
RT epithelial cells."
RL EMBO J. 12:641-650(1993).
CC -1- FUNCTION: IMPLICATED IN A NUMBER OF ADHERENCE FUNCTIONS. OPA
CC PROTEINS ARE IMPLICATED IN PATHOGENESIS AND ARE SUBJECT TO PHASE
CC VARIATION.

```

```

CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@sib.ac.uk).
CC EMBL: Z18933; CAA79366.1; -.
DR PIR: S28632; S28632. Porin opacity.
DR InterPro: IPR003394; Porin opacity.
DR Pfam: PF02462; Opacity; 1.
KW Outer membrane; Multigene family; Signal.
FT NON_TER 1 1
FT SIGNAL <1 1 POTENTIAL.
FT CHAIN 2 >234 OPACITY PROTEIN OPA55.
FT NON_TER 234 234
SQ SEQUENCE 234 AA; 26881 MW; 8EEB30B3A774C766 CRC64;

Alignment Scores:
Pred. No.: 14.6 Length: 234
Score: 82.50 Matches: 62
Percent Similarity: 31.10% Conservative: 31
Best Local Similarity: 20.74% Mismatches: 86
Query Match: 5.57% Indels: 120
DB: 1 Gaps: 14

US-10-062-624-41 (1-840) x OPAE_NEIGO (1-234)
QY 100 GATAACAAAGAGCTTCTACATTAGTCAAGTACAAATCAAGTATATCACACTTTAGA 159
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4 AspHisGlyArgGlyProTyrValGlnAlaAspLeuAlaTyrAlaTyrGluHisLeuThr 23
QY 160 AATCTCTGCTGAAGAACTCTTATTAAATGGAACAAATCTCTCACATAAAAGTTTTC 219
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
24 HisAspTyrProGluGlnThr----- 30
QY 220 GGACTAAAGAAAGATGGTGATATATAACAAACAAAGACGATTTTACAGAGTAGCT- 273
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
31 GlyThrLysLysAspLysIleSerThrValSerAspTyrPheArgAsnIleArgThrHis 50
QY 274 -----CCAGCATTTGATTTTCAAAATAACTTAATATCAGGATTTTTCAGGAAGTATT 324
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 SerIleHisProArgVal-----SerVal 58
QY 325 GGTACTCTATGGACGACCAAGATAGAACTGAA-----GCTGCATATCAACAATTTAAT 381
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
59 GlyTyrAspPheGlyGlyTyrArgIleAlaAspTyrAlaArgTyrArgLysTrpAsn 78
QY 382 CCA-----AAAAACACCGATAACAATGAT 405
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
79 AsnAsnLysTyrSerValAspIleLysGluLeuGlnAsnLysAsnGlnAsnLysArgAsp 98
QY 406 -----ACTGATATGCTGAATACTATAAACATTTTCATTTATCTCGTAAAGATGCAATG 459
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
99 LeuLysThrGluAsnGlnGluAsnGlyThrPheHisAlaValSer----- 113
QY 460 GAAGATCAGCAATATGATAGTACTTAAATAATGACGGCATAACTTTTATGTCTATTGATGTT 519
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
114 -----SerLeuGlyLeu 117
QY 520 AATACTTGTCTATGACATTACAGCTGAAGAGTATCTTTTGTACCATATGTCATGTCAGGT 579
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
118 SerAlaValTyrAspPheLysLeuAsnAsp---LysPheLysProTyr----- 132
QY 580 ATAGGACGACATCTTATCACTATTTTAAAGACCTCAATCTAAATTTTGTCTTACCAAGGA 639
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
133 IleGlyAla-----ArgValAlaTyr---Gly 140
QY 640 AAAATAGGTATTAGTTACCTTATCACCATATCACACCAAGAGTCTCTGCATTTATTTGGTATCTAC 699
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 141 HisValArgHisSerIleAspSerThrLysLysThrThrLysPheLeuThrSerSerTyr 160  
 QY 700 CATGGCGTTATGGTAATAAATTTGAGAAG----- 729  
 Db 161 GlyGlyLeuAsnProThrValTyrThrGluGluAsnThrGlnAsnAlaHisGlnSer 180  
 QY 730 -----ATACCTGTAATAACTCCTGTAGTATTAAATGATGCTCT 768  
 Db 181 AsnSerIleArgArgValGlyLeuGlyValIleAlaGlyValGlyPheAspIleThrPro 200  
 QY 769 CAACACCATCTGCTTCACTACTCTTACGCTGGATAC----- 807  
 Db 201 Lys-----LeuThrLeuAspThrGlyTyrArgTyrHisTyrTrpGlyArg 215  
 QY 808 -----TTGGCGGAGAAATGGAAATGAGTTCACTTC 840  
 Db 216 LeuGluAsnThrArgPheLysThrHisGluAlaSerLeuGlyValArgTyrArgPhe 234  
 RESULT 37  
 ROC1\_ARATH STANDARD; PRT; 289 AA.  
 AC Q9ZUU4;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Putative ribonucleoprotein At2G37220, chloroplast precursor.  
 GN AT2G37220 OR F3G5.1.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC euroside II; Brassicales; Brassicaceae; Arabidopsi.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X. Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,  
 RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,  
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,  
 RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,  
 RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,  
 RA Venter J.C.;  
 RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 thaliana."  
 RL Nature 402:761-768 (1999).  
 CC -!- FUNCTION: COULD BE INVOLVED IN SPLICING AND/OR PROCESSING OF  
 CHLOROPLAST RNA'S (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Chloroplast (Potential).  
 CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AC005896; AAC98043.1; --  
 DR HSP; P09651; LUPL.  
 DR SWISS-2DPAGE; Q9ZUU4; ARATH.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 2.  
 DR SMART; SM00360; RRM; 2.  
 DR PROSITE; PS0102; RRM; 2.  
 DR PROSITE; PS0030; RRM\_RNP\_1; 2.  
 KW RNA-binding; Ribonucleoprotein; Repeat; mRNA processing; Chloroplast;  
 KW Transit peptide.  
 FT TRANSIT 1 61 CHLOROPLAST (POTENTIAL).  
 FT CHAIN 62 289 PUTATIVE RIBONUCLEOPROTEIN AT2G37220.  
 FT DOMAIN 91 169 RNA-BINDING (RRM) 1.

FT DOMAIN 170 203 LINKER (GLY-RICH).  
 FT DOMAIN 204 282 RNA-BINDING (RRM) 2.  
 FT DOMAIN 168 171 SER-RICH.  
 SQ SEQUENCE 289 AA; 30718 MW; E500C3C0518369AD CRC64;  
 Alignment Scores:  
 Pred. No.: 14.8 Length: 289  
 Score: 82.50 Matches: 40  
 Percent Similarity: 39.46% Conservative: 18  
 Best Local Similarity: 27.21% Mismatches: 54  
 Query Match: 5.57% Indels: 35  
 DB: 1 Gaps: 6  
 US-10-062-624-41 (1-840) x ROC1\_ARATH (1-289)  
 QY 25 AGAAGCGGTTAATCTCATTAATCTTACCATATCATCTTTTGCAGATCCTGTA 84  
 Db 24 ArgProAlaSerValSerLeuSer-----ProSerLeuSerPhe----- 37  
 QY 85 GGTTCAAGAACTAATGATACAAAGAGGCTTCTACATTAGTGCAGAACTAATCCAAGT 144  
 Db 38 -----LysLeuAsnSerAspSerValSerPheSerIleAlaLysTyrAsnSerPro 55  
 QY 145 ATATCACACTTTAGAAAA-----TTCTCTGCTGAAGAA----- 177  
 Db 56 AlaSerArgPheAlaArgAsnValAlaIleThrSerGluPheGluValGluGluAspGly 75  
 QY 178 -----ACTCTTATTATGGAACAATCTCTCACTTAAAGTTTTCGGA 222  
 Db 76 PheAlaAspValAlaProProLysGluSerPheSerAlaAspLeuLysLeuPhe--- 94  
 QY 223 CTAAAGAAAGATGGTGATATACAAAAAAGACAGATTTTACAGAGTAGCTCCAGGCATT 282  
 Db 95 -----ValGlyAsnLeuProPheAsnValAspSerAlaGlnLeuAlaGlnLeuPhe 111  
 QY 283 GATTTCAAAATAACTTA-----ATATCAGGATTTTCA 315  
 Db 112 GluSerAlaGlyAsnValGluMetValGluValIleTyrAspLysIleThrGlyArgSer 131  
 QY 316 GGAAGTATGTTTACTCTTATCGAGCGACCAAGAAATAGAATTGAAGTGCATATCAACA 375  
 Db 132 ArgGlyPheGlyPheValThrMetSerSerValSerGluValGluAlaAlaGlnGln 151  
 QY 376 TTTAATCCAAAAACACCGAT 396  
 Db 152 PheAsnGlyTyrGluLeuAsp 158  
 RESULT 38  
 VACA\_HELPY  
 ID VACA\_HELPY STANDARD; PRT; 1290 AA.  
 AC P55981;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Vacuolating cytotoxin precursor.  
 GN VACA OR HP0887.  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 OC Helicobacter.  
 OX NCBI\_TaxID=210;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=26695 / ATCC 700392;  
 RX MEDLINE=97394467; PubMed=9252185;  
 RA Tomb J.-F., White O., Kerlavage A.R., Klenk H.-P., Gill S., Dougherty B.A.,  
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Kirkness E.F., Peterson S.,  
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Glodek A.,  
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Hickey E.K.,  
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,  
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,  
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,  
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,

RA Venter J.C.;  
 RT "The complete genome sequence of the gastric pathogen *Helicobacter pylori*."  
 RL Nature 388:539-547(1997).  
 CC -!- FUNCTION: INDUCES VACUOLATION OF EUKARYOTIC CELLS. CAUSES  
 CC ULCERATION AND GASTRIC LESIONS.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE000598; AAD07935.1; -.  
 DR TIGR; HP0887; -.  
 DR InterPro; IPR003842; Vaca.  
 DR Pfam; PF02691; Vaca; 1.  
 KW Cytotoxin; Toxin; Signal; Complete proteome.  
 FT SIGNAL 1 33 POTENTIAL.  
 FT CHAIN 34 ? VACUOLATING CYTOTOXIN.  
 FT PROPEP 7 1290 POTENTIAL.  
 SQ SEQUENCE 1290 AA; 139312 MW; F48B23513447A1AC CRC64;

## Alignment Scores:

Pred. No.: 16.3 Length: 1290  
 Score: 82.50 Matches: 49  
 Percent Similarity: 40.05% Conservative: 42  
 Best Local Similarity: 21.55% Mismatches: 73  
 Query Match: 5.57% Indels: 63  
 DB: 1 Gaps: 11

US-10-062-624-41 (1-840) x VACA\_HELPY (1-1290)

QY 1 ATGATTATAGAAATCTAGTAGAGCGGTTAATCTCATTA----- 45  
 Db 531 ValAenIleAsnLysLeuIleThrAlaSerThrAsnValAlaValLysAsnPheAsnIle 550  
 QY 46 -----ATGCAATCTTACCATATCATCTTTTGCAGAT 78  
 Db 551 AsnGluLeuValValLysThrAsnGlyValSerValGlyGluThrHisPheSerGlu 570  
 QY 79 CCTGTAGTTCAGAACTAATGATAAC----- 105  
 Db 571 AspileGlySerGlnSerArgIleAsnThrValArgLeuGluThrGlyThrArgSerIle 590  
 QY 106 -----AAGAAGCG-----TTCTACATT 123  
 Db 591 PheSerGlyValLysPheLysSerGlyGluLysLeuValIleAspGluPheTyTy 610  
 QY 124 AGT-----GCAAGTACAAATCCAGTATATACACTTTTAGAAATTC 165  
 Db 611 SerProTrpAsnTyPheAspAlaArgAsnIleLysAsnValGluIleThrArgLysPhe 630  
 QY 166 TCTGCTGAAGAACTCTTATTAAT-----GGAACAAATCTCTCACTAAAGATTTTC 219  
 Db 631 ---AlaSerSerThrProGluAsnProTrpGlyThrSerLysLeuMetPheAsnAsnLeu 649  
 QY 220 GGACTAAGAAGATGGTGATATACAAAAAGAGATTTTACAGAGTAGTCCAGGC 279  
 Db 650 ThrLeuGlyGlnAsnAla---ValMetAspTySerGlnPheSerAsnLeuThrIleGln 668  
 QY 280 ATTGATTTTCAAAATAACTTAATATATCAGGATTTTTCAGGAAGTATTGGTACTCTATGAC 339  
 Db 669 GlyAspPheIleAsnAsn-----GlnGlyThrIleAsnTyLeuValArg 683  
 QY 340 GGACCAAGATAGACTT-----GAAGCTGCATATCAACAACTTAATTCACAAA 387  
 Db 684 GlyGlyGlnValAlaThrLeuAsnValGlyAsnAlaAlaMetPhePhe----- 700  
 QY 388 AACCCGATACAAATGATGATGATAATGGTGAATGATCTATAAACAATTTTGCATTA---TCT 444

Db 701 ---SerAsnValAspSerAlaThrGlyPheTyGlnProLeuMetLysIleAsnSer 719  
 QY 445 CGTAAGATGCAATGGAAGATCAGCAATATCTAGTACTTAAATAATGACGGCATAACTTTT 504  
 Db 720 AlaglnAspLeuIleLysAsnLysGluHisValLeuLeuLysAlaLysIleGlyTy 739  
 QY 505 ATGTCATTGATGGTTAATACT 525  
 Db 740 GlyAsnValSerLeuGlyThr 746  
 RESULT 39  
 PURA\_DICDI  
 ID PURA\_DICDI STANDARD; PRT; 427 AA.  
 AC P21900;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Adenylosuccinate synthetase (EC 6.3.4.4) (IMP--aspartate ligase)  
 DE (Adss) (AMPSase).  
 GN PURA.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellum.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91115872; PubMed=1989999;  
 RA Wisemuller L., Wittbrodt J., Noegel A.A., Schleicher M.;  
 RT "Purification and cDNA-derived sequence of adenylosuccinate  
 RT synthetase from Dictyostelium discoideum."  
 RL J. Biol. Chem. 266:2480-2485(1991).  
 CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE DE NOVO PATHWAY OF PURINE  
 CC NUCLEOTIDE BIOSYNTHESIS.  
 CC -!- CATALYTIC ACTIVITY: Gtp + IMP + L-aspartate = GDP + phosphate +  
 CC adenylosuccinate.  
 CC -!- PATHWAY: AMP biosynthesis; first committed step.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC, AND PARTIALLY ASSOCIATED WITH  
 CC PARTICULATE FRACTIONS.  
 CC -!- DEVELOPMENTAL STAGE: SEEN DURING GROWTH BUT NOT DURING  
 CC DEVELOPMENT.  
 CC -!- SIMILARITY: BELONGS TO THE ADENYLOSUCCINATE SYNTHETASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M58471; AAA33167.1; -.  
 DR PIR; A38617; AJDODS.  
 DR HSSP; P12283; 1ADE.  
 DR DictyDb; DD05022; pura.  
 DR InterPro; IPR001114; Asucc\_synthetase.  
 DR Pfam; PF00709; Adenyloc synt; 1.  
 DR ProDom; PD001188; Asucc synthetase; 1.  
 DR TIGRfams; TIGR00184; pura; 1.  
 DR PROSITE; PS00513; ADENYLOSUCCIN SYN 2; 1.  
 DR PROSITE; PS01266; ADENYLOSUCCIN SYN 1; 1.  
 KW Purine biosynthesis; Ligase; GTP-binding.  
 FT NP\_BIND 11 17 GTP (POTENTIAL).  
 FT ACT\_SITE 143 143 BY SIMILARITY.  
 FT ACT\_SITE 150 150 BY SIMILARITY.  
 SQ SEQUENCE 427 AA; 47285 MW; D716A28D5D465767 CRC64;  
 Alignment Scores:  
 Pred. No.: 16.7 Length: 427  
 Score: 82.00 Matches: 50  
 Percent Similarity: 33.33% Conservative: 26  
 Best Local Similarity: 21.93% Mismatches: 84  
 Query Match: 5.54% Indels: 68  
 DB: 1 Gaps: 11





## Alignment Scores:

Pred. No.:	16.8	Length:	438
Score:	82.00	Matches:	33
Percent Similarity:	39.77%	Conservative:	37
Best Local Similarity:	18.75%	Mismatches:	59
Query Match:	5.70%	Indels:	47
DB:	1	Gaps:	7

US-10-062-624-41 (1-840) x INX7\_DROME (1-438)

```
QY      817 CTCGCCCAAGTATCAAGCTCAAGATTCTGTAAGCAGATGTGTTTTCAGGAGCATCAT 758
Db      182 ILeArgArGThrMeIleAspArgMetArgLeuAsnGlnSerTrp-----GlyAlaHis 199
QY      757 TTAATACTACAGGAG-----TTATTACAGGTA-----TCTTCT 725
Db      200 LeuValPheAlaGluValLeuAsnLeuLeuLeuLeuGlnIleThrTrpThrAsn 219
QY      724 CAAATTATTACCAATAACGCCATGGTAGTATCCACCAATAAATGCAGAGACTTCTGGTG 665
Db      220 -ArgPheLeuGlyGlnPheLeuThrLeuGlyProHisAlaLeuLysAsnArgTrpSe 239
QY      664 TCATAGG-----GTAACATAACCTATTTTCCTTGGTAAGCAAAATTTTAGATTGAGGT 611
Db      239 rAspGluLeuSerValLeuAspLeuValPheProLysIleThrLysCysLysPheHisLy 259
QY      610 CTTTAAATAAGTAGTATAGATCTGCTCTATACCTGCACATGTCATATGTTACGAAAGATA 551
Db      259 sPheGlyAspSerGlySerIleGlnMethHisAspAlaLeuCys----- 273
QY      550 CTCCTTCAGCTGTAATGTATAGCAAGTATTAAACCATCAATGACATAAAAGTTATGCCGT 491
Db      274 -----ValMetAlaLeuAsnIleMetAsnGluLysIleTyrIleIl 287
QY      490 CATTTTAAAGTACTACATATGCTGATCTTCATTCATGTCATCTTTACGAGATATGC AAAAT 431
Db      287 eLeuTrpPheTrpTyAlaPheLeuLeuIle----- 297
QY      430 GTTTATAGTATTCACCATTCATTCAGTATCATTTGTTATCGGTGTTTTTTGGATTAAATTCGT 371
Db      298 -----ValThrValLeuGlyLeuLeuTrpArgIleLeuTh 309
QY      370 GATATGCAGCTTCAAG-----TTCTATTCTTTGGTCCGTCAT 334
Db      309 rLeuCysPheTyrArgAsnValThrPheThrArgTrpSerLeuTyr 324
```

Search completed: July 8, 2003, 10:45:38  
Job time : 46.5 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 8, 2003, 10:39:56 ; Search time 87 Seconds  
(without alignments)  
3978.841 Million cell updates/sec

Title: US-10-062-624-41

Perfect score: 1481  
Sequence: 1 atgaattataagaataattc.....ttggaatgaggttcaccttc 840

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlp  
-O=/cgn2.1/USPTO/spool/US10062624/runat 08072003 093519 2013/app\_query.fasta\_1.1031  
-DB=SPTREMBL 21 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THRM SCORE=spt -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HSPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10062624 @CGN 1.1.138 @runat 08072003 093519 2013 -NCPU=6 -ICPU=3  
-NO.MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL.21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Match	Length	DB ID	Description
1	1462	98.7	280	2	Q9f476 ehrlichia c

SUMMARIES

ID	Q9f476	PRELIMINARY;	PRT;	280 AA.
AC	Q9f476;			
DT	01-MAR-2001 (TRENBLrel. 16, Created)			
DT	01-MAR-2001 (TRENBLrel. 16, Last sequence update)			
DT	01-JUN-2001 (TRENBLrel. 17, Last annotation update)			
DE	P28-2 (Major outer membrane protein P30-10).			
GN	P28-2 OR P30-10.			
OS	Ehrlichia canis.			
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;			
OC	Anaplasmataceae; Ehrlichia.			
OX	NCBI_TaxID=944;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=JAKE;			
RX	MEDLINE=99242757; PubMed=10255842;			
RA	McBride J.W., Yu, XJ, Walker D.H.;			
RT	"Molecular cloning of the gene for a conserved major immunoreactive 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic antigen."			
RL	Clin. Diagn. Lab. Immunol. 6:392-399(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=JAKE;			

# ALIGNMENTS

RESULT 1

ID	Q9f476	PRELIMINARY;	PRT;	280 AA.
AC	Q9f476;			
DT	01-MAR-2001 (TRENBLrel. 16, Created)			
DT	01-MAR-2001 (TRENBLrel. 16, Last sequence update)			
DT	01-JUN-2001 (TRENBLrel. 17, Last annotation update)			
DE	P28-2 (Major outer membrane protein P30-10).			
GN	P28-2 OR P30-10.			
OS	Ehrlichia canis.			
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;			
OC	Anaplasmataceae; Ehrlichia.			
OX	NCBI_TaxID=944;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=JAKE;			
RX	MEDLINE=99242757; PubMed=10255842;			
RA	McBride J.W., Yu, XJ, Walker D.H.;			
RT	"Molecular cloning of the gene for a conserved major immunoreactive 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic antigen."			
RL	Clin. Diagn. Lab. Immunol. 6:392-399(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=JAKE;			



```
QY 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAAGA-----ACTAATGATAACAAGAA 111
Db 21 TyrGlnSerPheAlaAspProValThrSerAsnAspThrGlyLeuAsnAspSerArgGlu 40
QY 112 GGCCTTCTACATTAGTGCACCAAGTACCAATCAAGTATATCACACTTTAGAAAATTTCTGTCT 171
Db 41 GlyPheTyrIleSerValLysTyrAsnProSerIleSerHisPheArgLysPheSerAla 60
QY 172 GAAGAACTCTTATTAATGGAACAATACTCTCACTAAAAAGTTTTCGGACTAAAGAA 231
Db 61 GluGluAlaProIleAsnGlyAsnThrSerIleThrLysLysValPheGlyLeuLysLys 80
QY 232 GATGGTGATATAACAAGAACAGATTTTACAGAGTAGTCTCCAGGCAATTCATTTCAA 291
Db 81 AspGlyAspIleAlaGlnSerAlaAsnPheAsnArgThrAspProAlaLeuGluPheGln 100
QY 292 AATAACTTAATATCAGGATTTTTCAGGAAGTATTTGTTACTCTATGCGGACCAAGATA 351
Db 101 AsnAsnLeuIleSerGlyPheSerGlySerIleGlyTyrAlaMetAspGlyProArgIle 120
QY 352 GAACCTTGAGCTGCATATCAACAATTTTAAATCCAAAAACACCGATACCAATGATCTGAT 411
Db 121 GluLeuGluAlaAlaTyrGlnLysPheAspAlaLysAsnProAspAsnAsnAspThrAsn 140
QY 412 AATGGTGAATCTATTAAGTATCTTCTGATACCATATGATGATGATGATGATGATGAT 471
Db 141 SerGlyAspTyrTyrLysTyrPheGlyLeuSerArgGluAspAlaIleAlaAspLysLys 160
QY 472 TATGTAGTACTTAAATGACGAGTATCTTCTGATACCATATGATGATGATGATGATGAT 531
Db 161 TyrValValLeuLysAsnGluGlyIleThrPheMetSerLeuMetValAsnThrCysTyr 180
QY 532 GACATTACAGCTGAAGGATGATCTTCTGATACCATATGATGATGATGATGATGATGAT 591
Db 181 AspIleThrAlaGluGlyValProPheIleProTyrAlaCysAlaGlyValGlyAlaAsp 200
QY 592 CTTATCACTATTTTAAAGACTCACTCAATTTTCTGATGATGATGATGATGATGATGAT 651
Db 201 LeuIleAsnValPheLysAspPheAsnLeuLysPheSerTyrGlnGlyLysIleGlyIle 220
QY 652 AGTTACCTCTACACAGAGTCTCTGATTTTATGATGATGATGATGATGATGATGAT 711
Db 221 SerTyrProIleThrProGluValSerAlaPheIleGlyTyrTyrHisGlyValIle 240
QY 712 GGTAAATTAATGAGAATACCTGTAATACTCTGTAGTATTAATGATGATGATGATGAT 771
Db 241 GlyAsnAsnPheAsnLysIleProValIleThrProValValLeuGluGlyAlaProGln 260
QY 772 ACCACATCTGCTCAGTACTCTGACGTTGATGATGATGATGATGATGATGATGATGAT 831
Db 261 ThrThrSerAlaLeuValThrIleAspThrGlyTyrPheGlyGlyGlyGlyValArg 280
QY 832 TTCACCTTC 840
Db 281 PheThrPhe 283

RESULT 3
Q9R443 PRELIMINARY; PRT; 282 AA.
AC Q9R443;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Major antigenic protein 1 like protein.
OS Cowdria ruminantium
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ANTIGUA, WELGEVONDEN, LEMCOT3, AND UMBANEIN;
RX MEDLINE=99216274; PubMed=10198207;
RA Sulisoma C.R., Mahan S.M., Barbet A.F.;
```

```
RT "The map1 Gene of Cowdria ruminantium Is a Member of a Multigene
RL Family Containing Both Conserved and Variable Genes.";
Biochem. Biophys. Res. Commun. 257:300-305(1999).
DR EMBL; AF125279; AAD26353.1; -
DR EMBL; AF125274; AAD26343.1; -
DR EMBL; AF125277; AAD26349.1; -
DR EMBL; AF125278; AAD26351.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 282 AA; 31040 MW; 0CB3C655775CBC90 CRC64;
```

## Alignment Scores:

```
Pred. NO.: 3.18e-84 Length: 282
Score: 1118.00 Matches: 215
Percent Similarity: 85.16% Conservative: 26
Best Local Similarity: 75.97% Mismatches: 38
Query Match: 75.49% Indels: 4
DB: 2 Gaps: 3
```

US-10-062-624-41 (1-840) x Q9R443 (1-282)

```
QY 1 ATGAATTATAAGAAAATTTCTAGTAAGAGCGCGTTAATCTCATTAAATGTCAATCTTACCA 60
Db 1 MetAsnTyrLysLysIleLeuValArgSerAlaLeuIleSerLeuMetSerPheLeuPro 20
QY 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAAGAACT-----AATGATAAC---AAAGAA 111
Db 21 TyrGlnSerPheAlaGluProValSerSerAsnAsnIleGlyAsnGluAlaLysGlu 40
QY 112 GGCCTTCTACATTAGTGCACCAAGTACCAATCCAAAGTATATCACACTTTAGAAAATTTCTGTCT 171
Db 41 GlyPheTyrIleSerAlaLysTyrAsnProSerIleProHisPheArgLysPheSerAla 60
QY 172 GAAGAACTCTTATTAATGGAACAATACTCTCACTAAAAAGTTTTCGGACTAAAGAA 231
Db 61 GluGluThrProValTyrGlyLysAspSerProThrLysLysValPheGlyLeuLysLys 80
QY 232 GATGGTGATATAACAAGAACAGATTTTACAGAGTAGTCTCCAGGCAATTCATTTCAA 291
Db 81 GluGlySerIleThrLysTyrSerAspPheThrArgThrAspIleSerPheGluGlyGln 100
QY 292 AATACTTAATATCAGGATTTTTCAGGAAGTATTTGTTACTCTATGCGGACCAAGATA 351
Db 101 AsnAsnPheIleSerGlyPheSerGlySerIleGlyTyrIleMetAspGlyProArgVal 120
QY 352 GAACCTTGAGCTGCATATCAACAATTTTAAATCCAAAAACACCGATACCAATGATCTGAT 411
Db 121 GluIleGluAlaLysTyrGlnLysPheAsnProLysAsn---ProAlaAsnGluThrAsp 139
QY 412 AATGGTGAATCTATTAAGCACTTTTTCGATTTATCTCGTAAAGATGCAATGGAGATCAGCA 471
Db 140 ThrSerAspTyrTyrLysHisTyrGlyLeuSerArgAlaGluAlaMetAlaAspLysLys 159
QY 472 TATGTAGTACTTAAATGACGCAATCTTTATGTCATTTATGTCATTTATGTCATTTATGTCAT 531
Db 160 TyrValValLeuThrAsnAsnGlyValThrPheSerSerLeuMetPheAsnAlaCysTyr 179
QY 532 GACATTACAGCTGAAGGATGATCTTTCGTACCATATGATGATGATGATGATGATGATGAT 591
Db 180 AspIleThrAlaGluGlyValProPheIleProTyrAlaCysAlaGlyIleGlyAlaAsp 199
QY 592 CTTATCACTATTTTAAAGACCTCAATCTAAATTTTCTTACCAAGGAAAAATAGGTATT 651
Db 200 LeuIleSerIlePheAspAspIleAsnLeuLysPheAlaTyrGlnGlyLysIleGlyIle 219
QY 652 AGTTACCTCTACACAGAGTCTCTGCAATTTATGTCGTCGTCGTCGTCGTCGTCGTCGTC 711
Db 220 SerTyrProIleThrProGluIleSerAlaPheIleGlyTyrTyrHisGlyValIle 239
QY 712 GGTAAATTAATGAGAATACCTGTAATACTCTGTCGTAGTATTAATGATGATGATGATGAT 771
Db 240 GlyAsnLysTyrAsnLysValProValLysLeuProValThrLeuThrAspAlaProGln 259
```



```

Db      21  TyrGlnSerPheAlaGluProValSerSerAsnAsnIleGlyAsnGluAsnAlaLysGlu 40
QY      112  GCCTTCTACATTAGTGCAGAACATCAAGTATATACACACTTTAGAAAATCTCTGCT 171
Db      41  GlyPheTyrIleSerAlaLysTyrAsnProSerIleProHisPheArgLysPheSerAla 60
QY      172  GAAGAACTCCTATTAAAGCAAAATCTCTCACTAAAGATTTTCGGACTAAAGAAA 231
Db      61  GluGluThrProValTyrGlyLysAspSerProThrLysLysValPheGlyLeuLysLys 80
QY      232  GATGCTGATATAACAAAAAGACCATTTTACAGAGTAGCTCCAGGCAATGATTTCAA 291
Db      81  AspGlySerIleThrLysTyrSerAspPheThrArgThrAspIleSerPheGluGlyGln 100
QY      292  AATACTTAATACAGATTTTCAGAGATATGTGTTACTCTATGACGCGACCAAGAATA 351
Db      101  SerAsnPheIleSerGlyPheSerGlySerIleGlyTyrIleMetAspGlyProArgVal 120
QY      352  GAACCTTGAGCTGCATATCACAATTTAATCCAAAAACACCGATAACAATGATCTGAT 411
Db      121  GluIleGluAlaAlaTyrGlnLysPheAsnProLysAsn---ProAlaAsnGluThrAsp 139
QY      412  AATGCTGAATACTATAAAACATTTTCGATCTCTGTAAGATGCAATGGAAGATCAGCAA 471
Db      140  ThrSerAspTyrTyrLysHisTyrGlyLeuSerArgAlaGluThrMetThrAspLysLys 159
QY      472  TATGTAAGTACTAAATGACGGCATAACTTTTATGTCATGATGGTTAAATCTGCTAT 531
Db      160  TyrValValLeuThrAsnAsnGlyValThrPheSerSerLeuMetPheAsnAlaCysTyr 179
QY      532  GACATTACAGCTGAAGGATATCTTCGTACCATATGTCATGTCAGGATATAGGACGAT 591
Db      180  AspIleThrAlaGluGlyValProPheIleProTyrAlaCysAlaGlyIleGlyAlaAsp 199
QY      592  CTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTCAAGAGAAAATAGGTATT 651
Db      200  LeuIleSerIlePheAspIleAsnLeuLysPheAlaTyrGlnGlyLysIleGlyIle 219
QY      652  AGTTACCTATCACACCAAGATCTCTGCATTTATGTTGGTATACCATCGGCTTATT 711
Db      220  SerTyrProIleThrProGluIleSerAlaPheIleGlyGlyTyrTyrHisGlyValle 239
QY      712  GGTAATAATTGAGAACATCCTGTA 738
Db      240  GlyAsnLysTyrAsnLysIleProVal 248

```

## RESULT 6

```

Q93DD4      .PRELIMINARY;      PRT;      275 AA.
AC      Q93DD4;
DT      01-DEC-2001 (TrEMBLrel. 19, Created)
DT      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE      Outer membrane protein p28.
OS      Escherichia chaffeensis.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC      Anaplasmataceae; Ehrlichia.
OX      NCBI_TaxID=945;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=V2;
RA      Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT      "Allele variation and patterns of transcription of the Ehrlichia
RT      chaffeensis 28 kDa outer membrane protein multigene family.";
RL      Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF393389; AAL2919.1; -
DR      InterPro; IPR002566; Surface_Ag_msp4.
DR      Pfam; PF01617; Surface_Ag_2; 1.
SQ      SEQUENCE 275 AA; 29974 MW; 2ECCF2F98B2E9D9 CRC64;

```

## Alignment Scores:

```

Pred. No.:      7.38e-47      Length:      275
Score:          665.50      Matches:      140

```

```

Percent Similarity:      63.99%      Conservative:      43
Best Local Similarity:      48.95%      Mismatches:      86
Query Match:      44.94%      Indels:      17
DB:      2      Gaps:      5

```

US-10-062-624-41 (1-840) x Q93DD4 (1-275)

```

QY      1  ATGAATTATAAGAAAATTTCTAGTAAGAGCCGGTTAAATCTCATTAAATGCTCAATCTTACCA 60
Db      1  MetAsnTyrLysLysValPheIleThrSerAlaLeuIleSerLeuIleSerSerLeuPro 20
QY      61  TATCAGCTTTTTCAGATCCTGTAGTTCAGAACTAATGATATACAAAGAGGCTTCTTAC 120
Db      21  GlyValSerPheSerAspProThrGlySerGlyIleAsnGlyAsn-----PheTyr 37
QY      121  ATTAGTGCAAAAGTACAATCCAAATGATATACACATTTTACAAAATTTCTCTGCTGAAGAACT 180
Db      38  IleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu--- 56
QY      181  CCTATTAAATGAACAAATTTCTCTCACTAAAGATTTTCGACTAAAGAAAGATGGTAT 240
Db      57  -----GluArgAsnThrThrAlaGlyValPheGlyLeuLysGlnAspTrpAsp 72
QY      241  -----ATACAAAAAAGACCATTTTACAGAGTAGCTCCAGGCATT 282
Db      73  GlySerAlaIleProHisThrGlnSerSerThrProPheThrValSerAsnTyrSerPhe 92
QY      283  GATTTTCAAAATACTTAATATACAGATTTTTCAGGAAGTATTTGGTTACTCTCTATGGACGGA 342
Db      93  LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetAspGly 112
QY      343  CCAAGATAGAACTTGAGCTGCATATCAACAATTTTAAATCCAAAAACACCGATAACAAT 402
Db      113  ProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn 132
QY      403  GATACGTAAATGCTGATACTATAACATTTTTCATTTTCATTTTCGTAAGATGCAATGAA 462
Db      133  -----TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerArgGlnAlaThrIleAla 150
QY      463  GATCAGCAATATGTAGTACTTTAAATAATCAGCGCATAACTTTTATGTCATTGATGTTAAT 522
Db      151  AspAsnLysPheValPheLeuLysAsnGluGlyLeuLeuAspIleSerPheMetLeuAsn 170
QY      523  ACTTGCTATGATACATTACAGCTGAAGAGTATCTTTTCGTACCATATGCAATGTCAGGTATA 582
Db      171  AlaCysTyrAspValIleSerGluGlyIleProPheSerProTyrIleCysAlaGlyIle 190
QY      583  GGACGACATCTTATCCTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGAAAA 642
Db      191  GlyThrAspLeuValSerMetPheGluValThrAsnProLysIleSerTyrGlnGlyLys 210
QY      643  ATAGTATTAGTTACCTTATCACACAGAGTCTCTGCATTTTATTTGGTGGATACCTACCAT 702
Db      211  LeuGlyLeuSerTyrSerIleSerProGluThrSerValPheValGlyGlyHisPheHis 230
QY      703  GCGCTATTGTTAATAAATTTGAGAAGATACCTCTGTAATCTCTGTAGTATTAAATGAT 762
Db      231  LysValMetGlyAsnGluPheArgAspIleProAlaIleIleProSerGlySerAsn--- 249
QY      763  GCTCTCAACACCATCTGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 822
Db      250  LeuThrGlyAsnHisProAlaIleValThrLeuAspValCysHisPheGlyIleGluLeu 269
QY      823  GGAATGAGGTTTCACTTCC 840
Db      270  GlyGlyArgPheAlaPhe 275

```

## RESULT 7

```

O85817      .PRELIMINARY;      PRT;      276 AA.
AC      O85817;
DT      01-NOV-1998 (TrEMBLrel. 08, Created)
DT      01-NOV-1998 (TrEMBLrel. 08, Last sequence update)

```

```

DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=SAPULPA;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=SAPULPA;
RA Yu X.-J., Walker D.H.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=V9;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF077734; AAC31547.1; -.
DR EMBL; AF393395; AAL12925.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 276 AA; 30027 MW; 2FD3698FC1F60BE CRC64;

Alignment Scores:
Pred. No.: 1.16e-45 Length: 276
Score: 651.00 Matches: 140
Percent Similarity: 63.4% Conservative: 44
Best Local Similarity: 48.2% Mismatches: 82
Query Match: 43.9% Indels: 24
DB: 2 Gaps: 7

US-10-062-624-41 (1-840) x 085817 (1-276)

QY 1 ATGAATTATAAGAAATCTAGTAAAGACGGCTTAATCTCAATGTCATCTTACCA 60
DB 1 MetAsnTyrLysValPheIleThrSerAlaLeuIleSerLeuIleSerSerLeuPro 20

QY 61 TATCAGTCTTTGTCAGATCTCTAGGTTCAAGAACTAATGATGATAACAAGAGCTTCTAC 120
DB 21 GlyValSerPheSerAspProThrGlySerGlyIleAsnGlyAsn-----PheTyr 37

QY 121 ATTAGTGCAAGTACATCAAGTATATACACTTTAGAAATTTCTCTGCTGAAGAACT 180
DB 38 IleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu--- 56

QY 181 CCTATTATGGAACAAATCTCTCACTAAAGTTTCGGACTAAAGAAAGATGGTGTAT 240
DB 57 -----GluArgAsnThrThrAlaGlyValPheGlyLeuLysGlnAspTrpAsp 72

QY 241 -----ATACAAAAAAGACGATTTTACAGAGTAGCTCCAGGCATT 282
DB 73 GlySerAlaIleProHisThrHisSerAsnThrProPheThrValSerAsnTyrSerPhe 92

QY 283 GATTTTCAAAATACTAATATATCAGGATTTTCAGGAGTATTGGTTACTCTATGACGGA 342
DB 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetAspGly 112

QY 343 CCAAGAATAAGACTTGAAGCTGCATATCAACAATTTAATCCAAAAACCCGATAACAT 402
DB 113 ProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn 132

QY 403 GATACCTGATAATGGTGATCTATAACATTTTGCATTATCTCGTAAAGATGCAATG--- 459
DB 133 -----TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerArgGlnAlaThrIleAla 150

```

```

QY 460 GAAGATCAGCAATATAGTAGTACTTAATAAATGACGGCATAACTTTTATGTCATTGATGTT 519
DB 151 AlaAspAsnThrPheValPheLeuLysAsnGluGlyLeuLeuAspIleSerPheMetLeu 170
QY 520 AATACCTGCTATGACATFACAGCTGAAGGAGTATCTTTTCGTACCATATGTCATGTCAGGT 579
DB 171 AsnAlaCysTyrAspValIleSerGluGlyIleProPheSerProTyrIleCysAlaGly 190
QY 580 ATAGAGCAGATCTTATCATTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGA 639
DB 191 IleGlyThrAspLeuValSerMetPheGluValThrAsnProLysIleSerTyrGlnGly 210
QY 640 AATATAGTATTAGTATACCTTATCACACCAAGTCTCTGTCATTTATTTGGTGGATAC 699
DB 211 LysLeuGlyLeuSerTyrSerIleSerProGluThrSerValPheValGlyGlyHisPhe 230
QY 700 CATGGCGTTATTGGTAAATAAATTTGAGAAGATACCTGTAATAACTCTT-----GTA 750
DB 231 HisLysValMetGlyAsnGluPheArgAspIleProAlaIleIleProSerGlySerSer 250
QY 751 GTATTAAATAGTCTCTCAAAACCATCTCTCTCAGTAACTCTTTCAGCTTGGATACCTTT 810
DB 251 LeuThrGlyAsnHisPro-----AlaIleValThrLeuAspValCysHisPhe 266
QY 811 GCGGAGGAAATGGAAATGAGGCTTACCTTC 840
DB 267 GlyIleGluLeuGlyGlyArgPheAlaPhe 276

RESULT 8
Q9ZGM9 PRELIMINARY; PRT; 280 AA.
ID Q9ZGM9
AC Q9ZGM9;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=JAX;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=JAX;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF077733; AAC31546.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30304 MW; 91C54AC7851B77F2 CRC64;

Alignment Scores:
Pred. No.: 2.06e-45 Length: 280
Score: 648.00 Matches: 138
Percent Similarity: 62.54% Conservative: 44
Best Local Similarity: 47.42% Mismatches: 87
Query Match: 43.75% Indels: 22
DB: 2 Gaps: 6

US-10-062-624-41 (1-840) x Q9ZGM9 (1-280)

QY 1 ATGAATTATAAGAAATTTCTAGTAAAGACGGCTTAATCTCAATGTCATCTTACCA 60
DB 1 MetAsnTyrLysValPheIleThrSerAlaLeuIleSerLeuIleSerSerLeuPro 20

```



Qy	61	TATCAGCTTTTTCGAGATCCCTGTAGTTC	CAAGAACTAATGATAACAAGAGCGCTTCAC	120
Db	21	GlyValSerPheSerAspProThrGlySerGlyIleAsnGlyAsn	-----PheTyr	37
Qy	121	ATTAGTGCAAAGTACAATCCCAAGATATATACACACTTTAGAAAAATCTCTGCTCAAGAAACT	180	
Db	38	IleSerGlyIleTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu	---	56
Qy	181	CCTATTAAATGGAACAAATTCCTCTCACTAAAAAGTTTTCGGGACTTAAAGAAAGATGGTGTAT	240	
Db	57	-----GluArgSerThrThrAlaGlyValPheGlyLeuLysGlnAspTrpAsp	72	
Qy	241	-----ATAACAAAAAAGACGATTTTACAAGAGTAGTCCAGGCATT	282	
Db	73	GlySerAlaIleSerHisThrThrProGluAsnIlePheThrValSerAsnTyrSerPhe	92	
Qy	283	GATTTTCAAAATAACTTAATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGACGCGA	342	
Db	93	LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetAspGly	112	
Qy	343	CCAGAATAGAACTTGAAGCTGCATATCAACAAATTAATCCAAAAACACCGATTAACAAAT	402	
Db	113	ProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnSer	132	
Qy	403	GATACTGATAATGGTGAATACAT	-----AAACATTTTGCATTATCTCGTAAAGAT	453
Db	133	TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerHisHisSerSerGlyThrSerMet	152	
Qy	454	GCAATGGAAGATCAGCAATATGTAGTACTTTAAAAATGACGGCATAACTTTTATGTCATTG	513	
Db	153	SerSerAlaGlyAspLysPheValPheLeuLysAsnGluGlyLeuLeuAspValSerPhe	172	
Qy	514	ATGGTTAATACTTGGTATGACATACACCTGAGGAGTATCTTTTCGTACCATATCATGT	573	
Db	173	MetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSerProTyrIleCys	192	
Qy	574	GCAGGTATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAACTTAAAAATTCGCTTAC	633	
Db	193	ValGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSerTyr	212	
Qy	634	CAAGGAAAAATAGTATTAGTTACCTATACACACAGAACTCTCTGCATTTATTCGTGGA	693	
Db	213	GlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluThrSerValPheIleGlyGly	232	
Qy	694	TACTACCATGCGGTTATTGGTAATAAATTCGAGAAGATACCTGTAACTCCCTGTAGTA	753	
Db	233	HisPheHisLysValIleGlyAsnGluPheArgAspIleProAlaValIlePro	-----250	
Qy	754	TTAATGATGCTCTCAAAACACATCT	-----GCTTCAGTAACTCTTCGACGTTGGATC	807
Db	251	--AsnGlySerSerLeuThrGlyAsnHisProAlaIleValThrLeuAspValCysHis	269	
Qy	808	TTTCGGCGGAAAAATGGATGAGGTTTCACCTTC	840	
Db	270	PheGlyIleGluLeuGlyGlyArgPheAlaPhe	280	

## RESULT 9

ID	Q93DD1	PRELIMINARY;	PRT;	276 AA.
AC	Q93DD1;			
DT	01-DEC-2001 (TrEMBLrel. 19, Created)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)			
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)			
DE	Outer membrane protein p28.			
OS	Ehrlichia chaffeensis.			
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;			
OC	Anaplasmataceae; Ehrlichia.			
OC	NCBI_TaxID=945;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=V7;			
RC	Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;			
RT	"Allele variation and patterns of transcription of the Ehrlichia			

```
QY 811 GCGGAGRAATTGGATGAGGTTTCACCTTC 840
Db 267 GlylleGluLeuGlyGlyArgPheAlaPhe 276

RESULT 10
Q93DD2 PRELIMINARY; PRT; 281 AA.
AC Q93DD2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V6;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RL chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF393392; AAL12922.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 281 AA; 30388 MW; B54DBB55AC839A9A CRC64;

Alignment Scores:
Pred. No.: 3,31e-45 Length: 281
Score: 645.50 Matches: 140
Percent Similarity: 62.20% Conservative: 41
Best Local Similarity: 48.11% Mismatches: 89
Query Match: 43.59% Indels: 21
DB: 2 Gaps: 6

US-10-062-624-41 (1-840) x Q93DD2 (1-281)

QY 1 ATGAATTATAAGAAATTTCTAGTAAGAGCGGTTTAATCTCATTAATGTCAATCTTACCA 60
Db 1 MetAsnTyrLysLysValPheIleThrSerAlaLeuLeuSerLeuLeuSerLeuPro 20
QY 61 TATCAGCTCTTTGCAGATCTCTGAGGTTCAAGAACTTAATGATAACAAGAGCTTCTAC 120
Db 21 GlyValSerPheSerAspProAlaGlySerGlyLeuAsnGlyAsn-----PheTyr 37
QY 121 ATTAGTCAAGTACCAATCAAGTATATCACACTTTAGAAAATTTCTCTGCAAGAACT 180
Db 38 IleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu--- 56
QY 181 CCTATTATGGAACAATTTCTCTCACTATAAAGTTTTCGAGCTAAAGAAAGTGTGAT 240
Db 57 -----GluArgAsnThrThrValGlyValPheGlyLeuLysGlnAsnTrpAsp 72
QY 241 -----ATAACAAAAGAGCGATTTTACAGAGAGTAGCTCCAGGCATT 282
Db 73 GlySerAlaIleSerAsnSerSerProAsnAspValPheThrValSerAsnTyrSerPhe 92
QY 283 GATTTTCAAAATAACTTAATATATCAGGATTTTTCAGGAAGTATGTTGTTACTCTATGGACGGA 342
Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetAspGly 112
QY 343 CCAGAAATAGAACTTGAAGTGCATATCAACAATTTAATCCAAAACACCGCATACAAT 402
Db 113 ProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn 132
QY 403 GATACTGATAATGGTCACTATCAACATTTTTCGTAATCTCTCGTAAGATGCAATGGAA 462
Db 133 -----TyrLysSerGluAlaHisArgTyrCysAlaLeuSerHisAsnSerAlaAlaAsp 150
QY 463 -----GATCAGCAATATGTAGTACTTTAAATAATGACGGCATAACTTTTATGTCA 510
Db 151 MetSerSerAlaSerAsnAsnPheValPheLeuLysAsnGluGlyLeuLeuAspIleSer 170
```

```
QY 511 TTGATGTTAATACTTCTGCTATGACATTACAGCTGAAGGAGTATCTTTCTGTTACCATATGCA 570
Db 171 PheMetLeuAsnAlaCysTyrAspValValGlyGlyGlyIleProPheSerProTyrIle 190
QY 571 TGTGAGGTATAGGACAGATCTTATCACTACTATTTTAAAGACCTCAATCTAAAATTTGCT 630
Db 191 CysAlaGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSer 210
QY 631 TACCAAGGAAAAATAGGTATTAGTATACCTATACACAGAGAGTCTTGCATTTATTTGCT 690
Db 211 TyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSerValPheIleGly 230
QY 691 GGATCTACCATCGGCTTATTGTTAATAAATTTGAGAGATACCTGTAACTAATCTCTGTA 750
Db 231 GlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrIleIleProThr 250
QY 751 GTATTAAATGATGCTCTCTCAA---ACCACATCTGCTTCAGTAACCTCTTTCAGCTTGATAC 807
Db 251 GlySerThrLeuAlaGluLysGlyAsnTyrProAlaIleValIleLeuAspValCysHis 270
QY 808 TTTGCGGAGAAAATTTGGAATGAGTTTCACTTCACCTTC 840
Db 271 PheGlyIleGluLeuGlyGlyArgPheAlaPhe 281

RESULT 11
Q9AC19 PRELIMINARY; PRT; 281 AA.
AC Q9AC19;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Major outer membrane protein P28.
GN P28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
RT are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=VI;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U72291; AAK28673.1; -
DR EMBL; AF393388; AAL12918.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 281 AA; 30343 MW; A99E5F7C4459AA9A CRC64;

Alignment Scores:
Pred. No.: 4e-45 Length: 281
Score: 644.50 Matches: 140
Percent Similarity: 62.20% Conservative: 41
Best Local Similarity: 48.11% Mismatches: 89
```

```

Query Match: 43.52% Indels: 21
DB: 2 Gaps: 6
US-10-062-624-41 (1-840) x Q9ACI9 (1-281)

QY 1 ATGAATTATAGAAATTCCTAGTAGAAGCGGTTAATCTCATTAATCTCAATCTTACCA 60
DB 1 MetAsnTyrLysValPheIleThrSerAlaLeuIleSerLeuIleSerLeuIlePro 20
QY 61 TATCAGTCTTTGTCAGATCCTGTAGTTCAAGAACTAATATGATAACAAGAGGCTTCTAC 120
DB 21 GlyValSerPheSerAspProAlaGlySerGlyIleAsnGlyAsn-----PheTyr 37
QY 121 ATTAGTGCAGAACTACAACTCCAGTATATACACATTTTACAAAATCTCTGCTGAAGAACT 180
DB 38 IleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu--- 56
QY 181 CCTATTATGGAACAAATCTCTCACTAAAGATTTTCGGACTAAAGAAAGAGTGTGAT 240
DB 57 -----GluArgAsnThrThrValGlyValPheGlyLeuLysGlnAsnTrpAsp 72
QY 241 -----ATAACAAAAGAACACGATTTTACAAGAGTAGCTCCAGGCATT 282
DB 73 GlySerAlaIleSerHisThrProGluAsnIlePheThrValSerAsnTyrSerPhe 92
QY 283 GATTTTCAAAATAACTTAAATATCAGGATTTTTCAGGAAGATTTGGTTACTTCTATGGACGGA 342
DB 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetAspGly 112
QY 343 CCAAGATAGAACTTGAAGCTGCATATCAACAATTTTAAATCCAAAACACCGATAACAAT 402
DB 113 ProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn 132
QY 403 GATACTGATAATGGTGAATCTATCAATATTTGCAATCTCGTAAAGATGCAATGAA 462
DB 133 -----TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerHisAsnSerAlaAlaAsp 150
QY 463 -----GATCAGCAATATAGTACTTATCACTATTTTAAAGACCTCAATCTAAATTTGCT 510
DB 151 MetSerSerAlaSerAsnAsnPheValPheLeuLysAsnGluGlyLeuLeuAspIleSer 170
QY 511 TTGATGGTTAATCTGTATGATACATACAGCTGAGGAGTATCTTTCGTACCATATGCA 570
DB 171 PheMetLeuAsnAlaCysTyrAspValValGlyGluGlyIleProPheSerProTyrIle 190
QY 571 TGTGAGGTATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCT 630
DB 191 CysAlaGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSer 210
QY 631 TACCAAGAAAATAGTATTAGTTACCTATCACACCAGAGTCTCTGCAATTTATGGT 690
DB 211 TyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSerValPheIleGly 230
QY 691 GGATACCTACCATGGGTTATGGTAAATTAATTTGAGAAGTACCTGTATTAACCTCTGTA 750
DB 231 GlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrIleIleProThr 250
QY 751 GTATTAAATGATGCTCCCTCAA---ACCACATCTGCTTCAGTAACCTCTGACGCTGGATAC 807
DB 251 GlySerThrLeuAlaGlyLysGlyAsnTyrProAlaIleValIleLeuAspValCysHis 270
QY 808 TTTCGGGAGAAATTTGGAATGAGGTTACCTTC 840
DB 271 PheGlyIleGluLeuGlyGlyArgPheAlaPhe 281

RESULT 12
O85816 PRELIMINARY; PRT; 280 AA.
AC O85816;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Outer membrane protein p28.

```

OS Ehrlichia chaffeensis.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=945;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=91HE17;  
RX MEDLINE=99175287; PubMed=10074538;  
RA Yu X.-J., McBride J.W., Walker D.H.;  
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in  
human isolates of Ehrlichia chaffeensis";  
RL J. Clin. Microbiol. 37:1137-1143 (1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=91HE17;  
RA Yu X.-J., Walker D.H.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=V8, AND V4;  
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;  
RT "Allele variation and patterns of transcription of the Ehrlichia  
chaffeensis 28 kDa outer membrane protein multigene family";  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF077732; AAC31545.1; -;  
DR EMBL; AF393394; AAL12924.1; -;  
DR EMBL; AF393390; AAL12920.1; -;  
DR InterPro; IPR002566; Surface\_Ag\_msp4.  
DR Pfam; PF01617; Surface\_Ag\_2; 1;  
SQ SEQUENCE 280 AA; 30277 MW; 91C54AC78507A63F CRC64;  
Alignment Scores:  
Pred. No.: 4,46-45 Length: 280  
Score: 644.00 Matches: 137  
Percent Similarity: 61.99% Conservative: 44  
Best Local Similarity: 46.92% Mismatches: 87  
Query Match: 43.48% Indels: 24  
DB: 2 Gaps: 6  
US-10-062-624-41 (1-840) x O85816 (1-280)  
QY 1 ATGAATTATAGAAATTCCTAGTAGAAGCGGTTAATCTCATTAATCTCAATCTTACCA 60  
DB 1 MetAsnTyrLysValPheIleThrSerAlaLeuIleSerLeuIleSerLeuIlePro 20  
QY 61 TATCAGTCTTTGTCAGATCCTGTAGTTCAAGAACTAATATGATAACAAGAGGCTTCTAC 120  
DB 21 GlyValSerPheSerAspProThrGlySerGlyIleAsnGlyAsn-----PheTyr 37  
QY 121 ATTAGTGCAGAACTACAACTCCAGTATATACACATTTTACAAAATCTCTGCTGAAGAACT 180  
DB 38 IleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu--- 56  
QY 181 CCTATTATGGAACAAATCTCTCACTAAAGATTTTCGGACTAAAGAAAGAGTGTGAT 240  
DB 57 -----GluArgSerThrThrAlaGlyValPheGlyLeuLysGlnAspTrpAsp 72  
QY 241 -----ATAACAAAAGAACACGATTTTACAAGAGTAGCTCCAGGCATT 282  
DB 73 GlySerAlaIleSerHisThrProGluAsnIlePheThrValSerAsnTyrSerPhe 92  
QY 283 GATTTTCAAAATAACTTAAATATCAGGATTTTTCAGGAAGATTTGGTTACTTCTATGGACGGA 342  
DB 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetAspGly 112  
QY 343 CCAAGATAGAACTTGAAGCTGCATATCAACAATTTTAAATCCAAAACACCGATAACAAT 402  
DB 113 ProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn 132  
QY 403 GATACTGATAATGGTGAATCTATCAATATTTGCAATCTCGTAAAGATGCAATGAA 462  
DB 133 -----TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerHisAsnSerAlaAlaAsp 150  
QY 463 -----GATCAGCAATATAGTACTTATCACTATTTTAAAGACCTCAATCTAAATTTGCT 510  
DB 151 MetSerSerAlaSerAsnAsnPheValPheLeuLysAsnGluGlyLeuLeuAspIleSer 170  
QY 511 TTGATGGTTAATCTGTATGATACATACAGCTGAGGAGTATCTTTCGTACCATATGCA 570  
DB 171 PheMetLeuAsnAlaCysTyrAspValValGlyGluGlyIleProPheSerProTyrIle 190  
QY 571 TGTGAGGTATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCT 630  
DB 191 CysAlaGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSer 210  
QY 631 TACCAAGAAAATAGTATTAGTTACCTATCACACCAGAGTCTCTGCAATTTATGGT 690  
DB 211 TyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSerValPheIleGly 230  
QY 691 GGATACCTACCATGGGTTATGGTAAATTAATTTGAGAAGTACCTGTATTAACCTCTGTA 750  
DB 231 GlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrIleIleProThr 250  
QY 751 GTATTAAATGATGCTCCCTCAA---ACCACATCTGCTTCAGTAACCTCTGACGCTGGATAC 807  
DB 251 GlySerThrLeuAlaGlyLysGlyAsnTyrProAlaIleValIleLeuAspValCysHis 270  
QY 808 TTTCGGGAGAAATTTGGAATGAGGTTACCTTC 840  
DB 271 PheGlyIleGluLeuGlyGlyArgPheAlaPhe 281

QY 454 GCAATGGAAGATCAGCAATATGTAGTACTTAAATAATGACGGCATCAACTTTTATGTGCTTG 513  
 Db 133 SerSerAlaGlyAspLysPheValPheLeuLysAsnGluGlyLeuLeuAspValSerPhe 172  
 QY 514 ATGGTTAATACCTATGACATATACAGCTGAAGAGTATCTTTCGTACCATATGCGATGT 573  
 Db 173 MetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSerProTyrIleCys 192  
 QY 574 GCAAGTATAGGAGCATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTAC 633  
 Db 193 ValGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSerTyr 212  
 QY 634 CAAGGAAATAAGTATTAGTTACCTATCACACAGAGTCTCGCATTTATTTGCTGA 693  
 Db 213 GlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluThrSerValPheIleGlyGly 232  
 QY 694 TACTACCATGGCGTATTGTTGTAATAATTTGAGAAGATACCTGTATTAATCTCT 747  
 Db 233 HisPheHisLysValIleGlyAsnGluPheArgAspIleProAlaValIleProSerGly 252  
 QY 748 ---GTAGTATTAAATGATCTCTCAACACCATCTGCTTCACTGATCACTTGAAGTGA 804  
 Db 253 SerSerLeuThrGlyAsnHisPro-----AlaIleValThrLeuAspValCys 268  
 QY 805 TACTTTGGCGGAGAAATGGAATGAGTTTCACCTTC 840  
 Db 269 HisPheGlyIleGluLeuGlyGlyArgPheAlaPhe 280

RESULT 13  
 ID O85358 PRELIMINARY; PRT; 276 AA.  
 AC O85358;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE 28 kDa major surface antigen-5 (Fragment).  
 OS Ehrlichia chaffeensis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Anaplasmataceae; Ehrlichia.  
 OX NCBI\_TaxID=945;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ARKANSAS;  
 RX MEDLINE=98321180; PubMed=9647746;  
 RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,  
 RA Alkham A.R.;  
 RT "Molecular characterization of a 28 kDa surface antigen gene family of  
 the tribe Ehrlichiae."  
 RL Biochem. Biophys. Res. Commun. 247:636-643(1998).  
 DR EMBL; AF062761; AAC26716.1; -  
 DR InterPro; IPR002566; Surface\_Ag\_msp4.  
 DR Pfam; PF01617; Surface\_Ag\_2; 1.  
 FT NON\_TER 276 276  
 SQ SEQUENCE 276 AA; 29782 MW; A9AB2A92263CA4EA CRC64;

## Alignment Scores:

Pred. No.: 6,93e-44 Length: 276  
 Score: 629.50 Matches: 137  
 Percent Similarity: 62.24% Conservative: 41  
 Best Local Similarity: 47.90% Mismatches: 87  
 Query Match: 21 Indels: 21  
 DB: 2 Gaps: 6

US-10-062-624-41 (1-840) x O85358 (1-276)

QY 1 ATGAATTATAAGAAATTTCTAGTAAGAAGCGGTTAATCTCAATATGTCATCTTACCA 60  
 Db 1 MetAsnTyrLysValPheIleThrSerAlaLeuIleSerLeuSerLeuPro 20  
 QY 61 TATCAGTCTTTGTCAGATCTCTGATGTTCAAGAACTTAATGATAACAAAGCGCTTCTAC 120  
 Db 21 GlyValSerPheSerAspProAlaGlySerGlyIleAsnGlyAsn-----PheTyr 37

QY 121 ATTAGTCAAAAGTACAATCCCAAGTATATCACACTTTTACAAAATTTCTCTGCTGAAGAACT 180  
 Db 38 IleSerGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu--- 56  
 QY 181 CCTATTATGAACACAAATTTCTCTCACTAAAAAAGTTTTTCGCACTAAAGAAAGATGGTAT 240  
 Db 57 -----GluTgAsnThrThrValGlyValPheGlyLeuLysGlnAsnTrpAsp 72  
 QY 241 -----ATAACAAAAAAGACGATTTTACAGAGTAGTCCCGCCATC 282  
 Db 73 GlySerAlaIleSerAsnSerSerProAsnAspValPheThrValSerAsnTyrSerPhe 92  
 QY 283 GATTTCAAAATAACTTAATATCAGGATTTTTCAGGAAGTATTGGTTACTCTATGACGCA 342  
 Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetAspGly 112  
 QY 343 CCAGAATAGAACCTTTGAAGCTGCATATCAACAATTTTAAATCCAAAAAACACCGATACAAT 402  
 Db 113 ProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn 132  
 QY 403 GATACCTGATTAATGTAATATCAATATTAACATTTTTCGATTTATCTCGTAAAGATGCAATGAA 462  
 Db 133 -----TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerHisAsnSerAlaAlaAsp 150  
 QY 463 -----GATCAGCAATATGCTAGTACTTAAATAATGACGGCATAACTTTTATGTCA 510  
 Db 151 MetSerSerAlaSerAsnAsnPheValPheLeuLysAsnGluGlyLeuLeuAspIleSer 170  
 QY 511 TTGATGGTTAATATCTGCTATGACATTCACAGCTGAAGAGTATCTTTTCGTACCATATGCA 570  
 Db 171 PheMetLeuAsnAlaCysTyrAspValValGlyGluGlyIleProPheSerProTyrIle 190  
 QY 571 TGTGAGTATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCT 630  
 Db 191 CysAlaGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSer 210  
 QY 631 TACCAAGGAAATAAGTATTAGTTTACCTATCACACCAAGAGTCTCTGCAATTTATGCT 690  
 Db 211 TyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSerValPheIleGly 230  
 QY 691 GGATACCTACCATGGCGTATTGTTGTAATAATTTGAGAAGATACCTGTATTAATCTCTGTA 750  
 Db 231 GlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProThrIleIleProThr 250  
 QY 751 GTATTAAATGATGCTCTCTCAA---ACCACATCTGCTTCAGTAACCTTCGACGTTGGATAC 807  
 Db 251 GlySerThrLeuAlaGlyLysGlyAsnTyrProAlaIleValIleLeuAspValCysHis 270  
 QY 808 TTTGGCGGAGAAATTTGGA 825  
 Db 271 PheGlyIleGluMetGly 276

RESULT 14  
 ID Q93DD3 PRELIMINARY; PRT; 280 AA.  
 AC Q93DD3;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Outer membrane protein p28.  
 OS Ehrlichia chaffeensis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Anaplasmataceae; Ehrlichia.  
 OX NCBI\_TaxID=945;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=V5;  
 RA Long S.W., Zhang X.-F., Qi H., Standart S., Walker D.H., Yu X.-J.;  
 RT "Allele variation and patterns of transcription of the Ehrlichia  
 chaffeensis 28 kDa outer membrane protein multigene family."  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF393391; AAL12921.1; -  
 DR InterPro; IPR002566; Surface\_Ag\_msp4.

DR Pfam: PF01617; Surface\_Ag\_2; 1.  
SQ SEQUENCE 280 AA; 30372 MW; C7BBC8710BC167E9 CRC64;

## Alignment Scores:

Pred. No.: 9,22e-44 Length: 280  
Score: 628.00 Matches: 134  
Percent Similarity: 60.96% Conservative: 44  
Best Local Similarity: 45.89% Mismatches: 90  
Query Match: 42.40% Indels: 24  
DB: 2 Gaps: 6

US-10-062-624-41 (1-840) x Q9ZGD3 (1-280)

```

Qy 1 ATGAATATAGAAAATCTAGTAAAGACGGCGTTAATCTCAATATGCTCAATCTTACCA 60
Db 1 MetAsnTyrLysValPheIleThrSerAlaLeuIleSerLeuSerLeuPro 20
Qy 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120
Db 21 GlyValSerPheSerAspProAlaGlyArgGlyLeuAsnGlyAsn-----PheTyr 37
Qy 121 ATTAGTCAAAAGTACAATCCAAGTATATCACACTTTAGAAAATCTCTGCTGAAGAACT 180
Db 38 IleArgGlyLysTyrMetProSerAlaSerHisPheGlyValPheSerAlaLysGlu--- 56
Qy 181 CCTATTATGAACAATACTCTCACTAAAGATTTTCGGACTAAAGAAAGATGGTGAT 240
Db 57 -----GluArgSerThrAlaGlyValPheGlyLeuLysGlnAspTyrAsp 72
Qy 241 -----ATACAAAAAAGACGATTTTACAAGTAGTACCTCCAGGCATT 282
Db 73 GlySerAlaIleSerHisThrProGluAsnIlePheThrValSerAsnTyrSerPhe 92
Qy 283 GATTTTCAAAATAAATAATACAGGATTTTCAGGAAGTATTGGTTACTCTATGACGGA 342
Db 93 LysTyrGluAsnThrProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetAspGly 112
Qy 343 CCAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAACACCGATAACAAT 402
Db 113 ProArgIleGluLeuGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnSer 132
Qy 403 GATACTGATATGCTGAATACAT-----AAACATTTTGCATTATCTCGTAAGAT 453
Db 133 TyrLysAsnGluAlaHisArgTyrCysAlaLeuSerHisHisSerSerGlyThrSerMet 152
Qy 454 GCAATGGAAGATCAGCAATATGTAGTACTTAAAAATGACGGCATAAATTTTATGTCATTG 513
Db 153 SerSerAlaGlyAspLysPheValPheLeuLysAsnGluGlyLeuLeuAspValSerPhe 172
Qy 514 ATGTTTAACTGCTATGACATACAGCTGAAGAGTATCTTTTCGTAACCATATGCAATG 573
Db 173 MetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSerProTyrIleCys 192
Qy 574 GCAGTATAGGACAGATCTTATCACTATTTTTAAAGACCTCAATCTAAATTTGCTTAC 633
Db 193 ValGlyIleGlyThrAspLeuValSerMetPheGluAlaThrAsnProLysIleSerTyr 212
Qy 634 CAAGAAAAATAGTATTAGTTACCTTATCACACAGAGTCTCTCGATTATTATTGGTGA 693
Db 213 GlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluThrSerValPheIleGlyGly 232
Qy 694 TACTACCATGCGTTATTGGTAATAATTTGAGAAGATACCTGTGAATACCTCT----- 747
Db 233 HisPheHisLysValIleGlyAsnGluPheArgAspIleProAlaValIleProSerGly 252
Qy 748 ---GTACTATTAAATGATGCTCTCAACACCATCTGCTTCAGTAACCTCTGAGCTTGA 804
Db 253 SerSerLeuThrGlyAsnHisPro-----AlaIleValThrLeuAspValCys 268
Qy 805 TACTTTGCGGAGAAATGGGAATGAGGTTCACCTTC 840
Db 269 HisPheGlyIleGluLeuGlyGlyArgPheAlaPhe 280

```

RESULT 15  
Q9ZGJ2

ID Q9ZGJ2 PRELIMINARY; PRT; 288 AA.  
AC Q9ZGJ2;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE 30-kDa major outer membrane protein (P28-8).  
GN P30 OR P28-8.  
OS Ehrlichia canis.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_taxid=944;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OKLAHOMA;  
RX MEDLINE=98371112; PubMed=9705412;  
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;  
RT "Cloning and characterization of multigenes encoding the  
RT immunodominant 30-kilodalton major outer membrane proteins of  
RT Ehrlichia canis and application of the recombinant protein for  
RT serodiagnosis." RT  
RL J. Clin. Microbiol. 36:2671-2680 (1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAKE;  
RX MEDLINE=99242757; PubMed=10225842;  
RA McBride J.W., Yu X.J., Walker D.H.;  
RT "Molecular cloning of the gene for a conserved major immunoreactive  
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic  
RT antigen." RT  
RL Clin. Diagn. Lab. Immunol. 6:392-399 (1999).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAKE;  
RX MEDLINE=20432107; PubMed=10974556;  
RA McBride J.W., Yu X.J., Walker D.H.;  
RT "A conserved, transcriptionally active p28 multigene locus of  
RT Ehrlichia canis." RT  
RL Gene 254:245-252 (2000).  
DR EMBL; AF078553; AAC68667.1; -  
DR EMBL; AF082744; AAG14362.1; -  
DR InterPro; IPR002566; Surface\_Ag\_msp4.  
DR Pfam; PF01617; Surface\_Ag\_2; 1.  
SQ SEQUENCE 288 AA; 31590 MW; 86DCAECB8E9BF5E CRC64;

## Alignment Scores:

Pred. No.: 3.49e-43 Length: 288  
Score: 621.00 Matches: 135  
Percent Similarity: 62.08% Conservative: 50  
Best Local Similarity: 45.30% Mismatches: 85  
Query Match: 41.93% Indels: 28  
DB: 2 Gaps: 7

US-10-062-624-41 (1-840) x Q9ZGJ2 (1-288)

```

Qy 1 ATGAATATAGAAAATCTAGTAAAGACGGCGTTAATCTCAATATGCTCAATCTTACCA 60
Db 1 MetAsnTyrLysValPheIleThrSerAlaLeuIleSerLeuMetSerPheLeuPro 20
Qy 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAAGAACTAATGATAACAAGAGGCG---TTC 117
Db 21 SerValSerPheSerGluSerIle-----HisGluAspAsnIleAsnGlyAsnPhe 37
Qy 118 TACATTAGTCCAAAGTACAATCCAAAGTATATCACACTTTAGAAAATCTCTGCTGAAGAA 177
Db 38 TyrIleSerAlaLysTyrMetProSerAlaSerHisPheGlyValPheSerValIysGlu 57
Qy 178 ACTCCTATTAAATGACAAATTTCTCTCACTAAAAAAGTTTTCGACTAAAAAGAGATGGT 237
Db 58 -----GluLysAsnThrThrGlyValPheGlyLeuLysGlnAspTyr 72
Qy 238 GAT---ATAACAAAAAAGACGAT-----TTT 261

```



```

Db      271 ValPheTyrPheGlyLeuGluGlyArgPheAsnPro 284
||||| ||||||| |||:||||| ||||||| |||
O52107 PRELIMINARY; PRT; 280 AA.
ID AC
AD O52107;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Major outer membrane protein OMP-1F.
GN OMP-1F.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
RT are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; U72291; AAC02940.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30731 MW; CCNA6C34E2AF393E CRC64;

Alignment Scores:
Pred. No.: 7.31e-42 Length: 280
Score: 605.00 Matches: 133
Percent Similarity: 63.39% Conservative: 54
Best Local Similarity: 45.08% Mismatches: 78
Query Match: 40.85% Indels: 30
DB: 2 Gaps: 10

US-10-062-624-41 (1-840) x O52107 (1-280)

Qy 1 ATGAATTATAGAAATCTAGTAAGAGCGGTTATCTCATTAAATGCTCAATCTTACCA 60
Db 1 MetAenCysLysPhePheileThrThrLeuValSerLeuMetSerPheLeuPro 20
Qy 61 TATCAGTCTTTTGAGATCCGTAGGTTCAAGAACTAATGATAACAAAGAGGC---TTC 117
Db 21 GlyleSerPheSerAspAlaVal-----GlnAsnAspAsnValGlyGlyAsnPro 37
Qy 118 TACATTAGTCAAAAGTACAAATCCAAAGTATATCACATTTTACAAAATCTCTGCTCAAGAA 177
Db 38 TyzileSerGlyLysTyrValProSerValSerHisPheGlyValPheSerAlaLysGln 57
Qy 178 ACTCCTATTATGACAAATCTCTCACTAAAAAGTTTCGCACTAAAGAAAGATGGT 237
Db 58 -----GluArgAsnThrThrThrGlyValPheGlyLeuLysGlnAspTrp 72
Qy 238 GAT-----ATAACAAAAAGAC---CATTTTCAAGAGTAGCTCCAGGCAATTCAT 285
Db 73 AspGlySerThrIleSerLysAsnSerProGluAsnThrPheAsnValProAsnTyrSer 92
Qy 286 TTT-----CAAAATAATTAATATCAGATTTTTCAGGAAGTATTTGGTTACTCTATGAC 339
Db 93 PheLysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaValGlyTyrLeuMetAsn 112
Qy 340 GGACCAAGATAGAACTTTGAAGCTGCATATCAACAATTTTAAATCCAAAAAACCCGATAAC 399

```

---

```

Db      113 GlyProArgIleGluLeuGluMetSerTyrGluThrPheAspValLysAsnGlnGlyAsn 132
||||| ||||||| ||||||| |||:||||| ||||||| |||
Qy 400 AATGATATCTGATAATGGTGAATACTATAAACTATTTTGCATTATCTCGT----- 447
Db 133 AsnTyrLysAsnAspAla-----HisLysTyrTyrAlaLeuThrHisAsnSerGlyGly 150
Qy 448 AAAGATCAATGGAAGATCAGCAATATGTAGTACTTTAAATAGAGCGCATAACTTTTATG 507
Db 151 LysLeuSerAsnAlaGlyAspLysPheValPheLeuLysAsnGluGlyLeuLeuAspIle 170
Qy 508 TCATTGATGGTTATATCTCTGATACATTACACTGAGGAGTATCTTTTGTACCATAT 567
Db 171 SerLeuMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSerProTyr 190
Qy 568 GCATGTGCAGGTATAGGAGCAGACTTATCACTACTATTTTAAAGACCTCAATCTTAAATTT 627
Db 191 IleCysAlaGlyValGlyThrAspLeuIleSerMetPheGluAlaIleAsnProLysIle 210
Qy 628 GCTTACCAAGAAAATAGGTATTAGTATACCTATCACACCAAGAGTCTCTGCATTAT 687
Db 211 SerTyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSerValPheVal 230
Qy 688 GGTGATACTACCATGGCGTATTGCTTAATAAATTTGAGAAGATACCTGTAACTACTCT 747
Db 231 GlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProAlaMetIlePro 250
Qy 748 GTAGTATTAAATGATGCTCTCAACACCATCTCTTCA-----GTAACCTCT 795
Db 251 -----SerThrSerThrLeuThrGlyAsnHisPheThrIleValThrLeu 265
Qy 796 GACGTGCTGATCTTTGGCGGAGAAATTTGGAATGAGTTTCACCTTC 840
Db 266 SerValCysHisPheGlyValGluLeuGlyGlyArgPheAsnPro 280

RESULT 18
O85357 PRELIMINARY; PRT; 280 AA.
ID AC
AC O85357;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 28 kDa major surface antigen-4.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulseona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RA Allemen A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
RT the tribe Ehrlichiae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
DR EMBL; AF062761; AAC26720.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30743 MW; FBB841DAF08EE4DC CRC64;

Alignment Scores:
Pred. No.: 1.07e-41 Length: 280
Score: 603.00 Matches: 133
Percent Similarity: 63.39% Conservative: 54
Best Local Similarity: 45.08% Mismatches: 78
Query Match: 40.72% Indels: 30
DB: 2 Gaps: 10

US-10-062-624-41 (1-840) x O85357 (1-280)

Qy 1 ATGAATTATAGAAATCTAGTAAGAGCGGTTATCTCATTAAATGCTCAATCTTACCA 60

```

Db 1 MetAsnCysLysPhePheIleThrThrLeuValSerLeuMetSerPheLeuPro 20  
QY 61 TATCAGTCTTTTGCAGACTCTAGTTCAGAACTAATGATAACAAGAGGC---TTC 117  
Db 21 GlyIleSerPheSerAspAlaVal-----GlnAsnAspValGlyGlyAsnPhe 37  
QY 118 TACATTAGTCAAGAGTACAACTCAAGTATATCACACTTTAGAAAATCTCTGCTCAAGAA 177  
Db 38 TyrIleSerGlyLysTyrValProSerValSerHisPheGlyValPheSerAlaLysGln 57  
QY 178 ACTCCTATTATGGAACAATCTCTCACTAAATAAGTTTCGGACTAAAGAAAGATGCT 237  
Db 58 -----GluArgAsnThrThrIleGlyValPheGlyLeuLysGlnAspTrp 72  
QY 238 GAT-----ATAACAAAAAGAC---GATTTTACAGAGTAGTCTCCAGGCATTCAT 285  
Db 73 AspGlySerThrIleSerLysAsnSerProGluAsnThrPheAsnValProAsnTyrSer 92  
QY 286 TTT-----CAAAATAACTTATATACAGATTTTCAGGAAGTATGTTTACTCTATGGAC 339  
Db 93 PheLysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaValGlyTyrLeuMetAsn 112  
QY 340 GCACCAAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGATAAC 399  
Db 113 GlyProArgIleGluLeuGluMetSerTyrGluThrPheAspValLysAsnGlnGlyAsn 132  
QY 400 AATGATCTGATAATGGTGAATACATATAACATTTTGGCATTTATCTCGT----- 447  
Db 133 AsnTyrLysAsnAspAla-----HisLysTyrTyrAlaLeuThrHisAsnSerGlyGly 150  
QY 448 AAGATGCAATGGAAGATCAGCAATATGATAGTACTTAAATAAGCGGATATCTTTATG 507  
Db 151 LysLeuSerAsnAlaGlyAspLysPheValPheLeuLysAsnGlyLeuLeuAspIle 170  
QY 508 TCATTGATGGTTAATCTTCTGATGACATTCAGCTGGAAGGAGTATCTTTCTGATCCATAT 567  
Db 171 SerLeuMetLeuAsnAlaCysTyrAspValIleSerGluGlyIleProPheSerProTyr 190  
QY 568 GCATGTCAGGTATAGGACAGATCTTATCACTATTTTAAAGACTCAATCTAAATTT 627  
Db 191 IleCysAlaGlyValGlyThrAspLeuIleSerMetPheGluAlaIleAsnProLysIle 210  
QY 628 GCTTACCAGGAAAAATAGTATTAGTTACCTTATCACCTATCACACAGAGTCTCTGCTATT 687  
Db 211 SerTyrGlnGlyLysLeuGlyLeuSerTyrSerIleSerProGluAlaSerValPheVal 230  
QY 688 GGTGATACTACCATGGCGTTATGGTATATAATTTGAAGATACCTGTATAACTCTCT 747  
Db 231 GlyGlyHisPheHisLysValIleGlyAsnGluPheArgAspIleProAlaMetIlePro 250  
QY 748 GTAGTATTAAATGATGCTCCTCAACACCATCTGCTTCA-----GTAACCTTT 795  
Db 251 -----SerThrSerThrLeuThrGlyAsnHisPheThrIleValThrLeu 265  
QY 796 GACGTTGGATCTTTGGCGGAGAAATTTGGAATGAGTTCCACCTTC 840  
Db 266 SerValCysHisPheGlyValGluLeuGlyGlyArgPheAsnPhe 280  
RESULT 19  
Q9ADV3 PRELIMINARY; PRT; 280 AA.  
AC Q9ADV3  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
DE Major outer membrane protein P30-2.  
GN P30-2.  
OS Ehrlichia canis.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=944;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=OKLAHOMA;  
RX MEDLINE=983771112; PubMed=9705412;  
RA Chao N., Unver A., Zhi N., Rikhisia Y.;  
RT "Cloning and characterization of multigenes encoding the  
RT immunodominant 30-kilodalton major outer membrane proteins of  
RT Ehrlichia canis and application of the recombinant protein for  
RT serodiagnosis.";  
RL J. Clin. Microbiol. 36:2671-2680 (1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OKLAHOMA;  
RX MEDLINE=21153566; PubMed=11254561;  
RA Chao N., Rikhisia Y., Unver A.;  
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer  
RT Membrane Protein Multigene Family in Ehrlichia canis and E.  
RT Chaffeensis.";  
RL Infect. Immun. 69:2083-2091 (2001).  
DR EMBL; AF078553; AAK28699.1; -  
DR InterPro; IPR002566; Surface\_Ag\_msp4.  
DR Pfam; PF01617; Surface\_Ag\_2; 1. 2723BBE1C7B68A91 CRC64;  
SQ SEQUENCE 280 AA; 30803 MW; 2723BBE1C7B68A91 CRC64;  
Alignment Scores:  
Pred. No.: 1.56e-41 Length: 280  
Score: 601.00 Matches: 129  
Percent Similarity: 61.59% Conservative: 49  
Best Local Similarity: 44.84% Mismatches: 93  
Query Match: 40.58% Indels: 18  
DB: 2 Gaps: 6  
US-10-062-624-41 (1-840) x Q9ADV3 (1-280)  
QY 1 ATGAATTATAGAAAAATCTAGTAGAAGCGGTTAATCTCAATTAATCTCAATCTTACCA 60  
Db 1 MetAsnCysLysLysIleLeuIleThrThrAlaLeuMetSerLeuMetTyrTyrAlaPro 20  
QY 61 TATCAGTCTTTTGCAGATCCGTAGGTTCAGAACTAATGATAACAAGAGGCTTCTAC 120  
Db 21 SerLeuSerPheSerAspThrIle-----GlnAspAspAsnThrGlySerPheTyr 37  
QY 121 ATTAGTCAAAAGTACAATCCAAAGTATATCACACTTTAGAAAAATCTCTGCTGAAGAACT 180  
Db 38 IleSerGlyLysTyrValProSerValSerHisPheGlyValPheSerAlaLysGlu--- 56  
QY 181 CCTATTATGAACAACAATCTCTCACTAAAAAGTTTCGAGACTAAAGAAAGAT----- 234  
Db 57 -----GluArgAsnSerThrValGlyValPheGlyLeuLysHisAspTrpAsn 72  
QY 235 --GGTGATATAACAAAA-----AAAGACGATTTTACAAGAGTAGTCTCCAGGCATT 282  
Db 73 GlyGlyThrIleSerAsnSerSerProGluAsnIlePheThrValGlnAsnTyrSerPhe 92  
QY 283 GATTTCAAAAATAACTTAATATCAGGATTTTTCAGGAAGTATTTGGTTACTCTATGGACGGA 342  
Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGlyGly 112  
QY 343 CCAGATAGAACTTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGATACCAAT 402  
Db 113 ProArgIleGluLeuGluValLeuTyrGluThrPheAspValLysAsnGlnAsnAsn 132  
QY 403 GATCTGATATGTTGTAATACTAT-----AAACATTTTTCGATTTATTCGTAAGAT 453  
Db 133 TyrLysAsnGlyAlaHisArgTyrCysAlaLeuSerHisHisSerSerAlaThrAsnMet 152  
QY 454 GCAATGGAAGATCAGCAATATAGTAGTACTTAAAAATAGCGCATAACTTTTATGTCATTG 513  
Db 153 SerSerAlaSerAsnLysPheValPheLeuLysAsnGluGlyLeuIleAspLeuSerPhe 172  
QY 514 ATGGTTAATACTTGTATGATACATTCACCTCAAGAGAGTATCTTTCGTACCATATGCTATG 573  
Db 173 MetIleAsnAlaCysTyrAspIleIleLeuGlyMetProPheSerProTyrIleCys 192  
QY 574 GCAGGTATAGGACGAGATCTTATCACTATTATTTTAAAGACCTCAATCTAAAAATTTGCTTAC 633



```
Db 193 AlaGlyValGlyThrAspValValSerMetPheGluAlaIleAsnProLySileSerTyr 212
Qy 634 CAAGGAAATAGGTATTAGTTACCTATCACACCAGAGTCTCTGCAATTTATTGGTGA 693
Db 213 GlnGlyLysLeuGlyLeuGlyTyrSerIleSerSerGluAlaSerValPheIleGlyGly 232
Qy 694 TACTACCATGGCTTATTGGTAATAATTTGAGAGATACCTGTAATAACTCTCTGTAGTA 753
Db 233 HisPheHisArgValIleGlyAsnGluPheArgAspIleProAlaMetValProSerGly 252
Qy 754 TTAATGATGCTCTCAACACACATCTCTCAGTAACCTCTGAGCTTGGATACCTTGGC 813
Db 253 SerAsn---LeuProGluAsnGlnPheAlaIleValThrLeuAsnValCysHisPheGly 271
Qy 814 GGAGAAATGGAATGAGGTTCCACCTTC 840
Db 272 LeuGluLeuGlyGlyArgPheAsnPhe 280

RESULT 20
Q9F473 PRELIMINARY; PRT; 280 AA.
AC Q9F473;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P28-6.
GN Ehrlichia canis.
OS Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAKE;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "Molecular Cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen."
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=JAKE;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis."
RL Gene 254:245-252(2000).
DR EMBL; AF082744; AAG14361.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30762 MW; BE284A4B94FE3123 CRC64;

Alignment Scores:
Pred. No.: 1.89e-41 Length: 280
Score: 600.00 Matches: 129
Percent Similarity: 61.59% Conservative: 49
Best Local Similarity: 44.64% Mismatches: 93
Query Match: 40.51% Indels: 18
DB: 2 Gaps: 6

US-10-062-624-41 (1-840) x Q9F473 (1-280)
Qy 1 ATGAATTATAAGAAATCTAGTAGAGCGGTTAATCTCATTAATGCTCAATCTTACCA 60
Db 1 MetAsnCysLysIleLeuIleThrAlaLeuMetSerLeuMetCysTyrAlaPro 20
Qy 61 TATCAGTCTTTGCAAGCTCTGTAGTTCAGAACTAATGATAACAAAGAGGCTTCTAC 120
Db 21 SerIleSerPheSerAspThrIle-----GlnAspAspAsnThrGlySerPheTyr 37
Qy 121 ATTAGTGCAAAGTACAATCCAAAGTATATCACTTTAGAAAAATCTCTGTGGAAGAACT 180
```

```
Db 38 IleSerGlyLysTyrValProSerValSerHisPheGlyValPheSerAlaLysGlu--- 56
Qy 181 CCTATTATGAACAAATTTCTCTCACTAAAAAGTTTTTCGAGCTAAAGAAAGAT----- 234
Db 57 -----GluArgAsnSerThrValGlyValPheGlyLeuLysHisAspTrpAsn 72
Qy 235 ---GGTGATATAACAAA-----AAAGACGATTTTACAAGAGTAGTCCAGGCATT 282
Db 73 GlyGlyThrIleSerAsnSerSerProGluAsnIlePheThrValGlnAsnTyr-SerPhe 92
Qy 283 GATTTTCAATTAATTAATATCAGGATTTTCAGGAAGTATTGGTTACTCTTATGGACCGA 342
Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGlyGly 112
Qy 343 CCAAGATAGAACTTGAAGCTGCATATCAACAAATTTTAAATCCAAAAAACACCGATAACAAT 402
Db 113 ProArgIleGluLeuGluValLeuTyrGluThrPheAspValLysAsnGlnAsnAsn 132
Qy 403 GATACTGATTAATGCTGAATACTAT-----AAACATTTTGCATTATCTCGTAAAGAT 453
Db 133 TyrLysAsnGlyAlaHisArgTyr-CysAlaLeuSerHisHisSerSerAlaThr-SerMet 152
Qy 454 GCATGCAAGATCAGCAATATGATAGTACTTAAATGACGCGCATAACTTTTATGTCATTG 513
Db 153 SerSerAlaSerAsnLysPheValPheLeuLysAsnGluGlyLeuIleAspLeuSerPhe 172
Qy 514 ATGTTTAATTAATCTGCTATGACATTACAGCTGAAGAGTATCTTTTCGTACCATATGTCATGT 573
Db 173 MetIleAsnAlaCysTyrAspIleIleGlyMetProPheSerProTyrIleCys 192
Qy 574 GCAGTATAGAGCAGATCTTATCATCTATTTTAAAGACCTCAATCTAAAAATTTGCTTAC 633
Db 193 AlaGlyValGlyThrAspValValSerMetPheGluAlaValAsnProLySileSerTyr 212
Qy 634 CAAGGAAATAGGTATTAGTTACCTATCACACCAGAGTCTCTGCAATTTATTGGTGA 693
Db 213 GlnGlyLysLeuGlyLeuGlyTyrSerIleSerSerGluAlaSerValPheIleGlyGly 232
Qy 694 TACTACCATGGCTTATTGGTAATAATTTGAGAGATACCTGTAATAACTCTCTGTAGTA 753
Db 233 HisPheHisArgValIleGlyAsnGluPheArgAspIleProAlaMetValProSerGly 252
Qy 754 TTAATGATGCTCTCAACACACATCTGTTTCAGTAACTCTTTCAGTTGGATACTTTGGC 813
Db 253 SerAsn---LeuProGluAsnGlnPheAlaIleValThrLeuAsnValCysHisPheGly 271
Qy 814 GGAGAAATGGAATGAGGTTCCACCTTC 840
Db 272 LeuGluLeuGlyGlyArgPheAsnPhe 280

RESULT 21
Q9L6V5 PRELIMINARY; PRT; 285 AA.
AC Q9L6V5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P29-9 (Major outer membrane protein OMP-1Y).
GN P28-9 OR OMP-1Y.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RA Yu X.J., McBride J.W., Zhang X., Walker D.H.;
RT "Characterization of the complete transcriptionally active Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family."
RL Gene 248:29-68(2000).
RN [2]
RP SEQUENCE FROM N.A.
```



QY 235 GGTGAT-----ATAACAAAAAGAGGATTTTACAGAGTAGTCCAGGC 279  
 Db |||||  
 QY 101 TrpAspGlySerProIleLeuLeuAsnLysHisAlaAspPheThrValProAsnTyrSer 120  
 Db |||||  
 QY 280 ATTGATTTTCAAAATAACTTAATATCAGCATTTTCAGGAAGTATTGGTTACTCTATGGAC 339  
 Db |||||  
 QY 121 PheArgTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGly 140  
 Db |||||  
 QY 340 GGACCAAGATAGACTTGAAGTGCATATCAACAATTTAATCCAAAAAACACCGATTAAC 399  
 Db |||||  
 QY 141 GlyProArgIleGluPheGluLeuSerTyrGluAlaPheAspValLysSerProAsnIle 160  
 Db |||||  
 QY 400 AATGATCTGATATCGGTGATATCAATATAAATTTTGCATTTATCTCCGTAAGATGCAATG 459  
 Db |||||  
 QY 161 AsnTyrGlnAsnAspAlaHisArgTyrCysAlaLeuSerHisHisThrSerAlaAlaMet 180  
 Db |||||  
 QY 460 GAAGATCAGCAATATGTAGTACTTAAATAATGACGGCATAACTTTTATGTCATTGATGTT 519  
 Db |||||  
 QY 181 GluAlaAspLysPheValPheLeuLysAsnGluGlyLeuIleAspIleSerLeuAlaIle 200  
 Db |||||  
 QY 520 AATACTTCTGATGATATCAGCTGAGGAGTATCTTTTCGTACCATATGCGATGTCAGGT 579  
 Db |||||  
 QY 201 AsnAlaCysTyrAspIleAsnAspLysValProValSerProTyrIleCysAlaGly 220  
 Db |||||  
 QY 580 ATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGA 639  
 Db |||||  
 QY 221 IleGlyThrAspLeuIleSerMetPheGluAlaThrSerProLysIleSerTyrGlnGly 240  
 Db |||||  
 QY 640 AAAATAGGTATTAGTTACCTATCACACGAGGTCCTGCGATTTTATTTGTTGGTACTAC 699  
 Db |||||  
 QY 241 LysLeuGlyIleSerTyrSerIleAsnProGluThrSerValPheIleGlyHisPhe 260  
 Db |||||  
 QY 700 CATGCGTATTGGTAAATAATTTGAGAAGATACCTGTAATACTCTCTGATGATTTAAAT 759  
 Db |||||  
 QY 261 HisArgIleIleGlyAsnGluPheArgAspIleProAlaIleValProSerAsnSerThr 280  
 Db |||||  
 QY 760 GATGCTCTCAACACCATCTGTTTCACTACTCTTGAGTCTTGAGTATTTGGGGAGAA 819  
 Db |||||  
 QY 281 ThrIleSerGlyProGlnPheAlaThrValThrLeuAsnValCysHisPheGlyLeuGlu 300  
 Db |||||  
 QY 820 ATTGGAATGAGTTCACCTTC 840  
 Db |||||  
 QY 301 LeuGlyGlyArgPheAsnPhe 307  
 Db |||||

## RESULT 23

Q9R8A9 PRELIMINARY; PRT; 278 AA.  
 AC Q9R8A9  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE 28 kDa outer membrane protein (Fragment).  
 OS Ehrlichia canis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Anaplasmataceae; Ehrlichia.  
 OX NCBI\_TaxID=944;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LOUISIANA;  
 RX MEDLINE=99242757; PubMed=10225842;  
 RA McBride J.W., Yu, X.J., Walker D.H.;  
 RT "Molecular cloning of the gene for a conserved major immunoreactive  
 RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic  
 RT antigen."  
 RL Clin. Diagn. Lab. Immunol. 6:392-399 (1999).  
 DR EMBL; AF082745; AAC64551.1; -;  
 DR InterPro; IPR002566; Surface\_Ag\_msp4.  
 DR Pfam; PF01617; Surface\_Ag\_2; 1.  
 FT NON TER 278  
 SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Alignment Scores:

Pred. No.: 5.8e-40 Length: 278

Score: 582.00 Matches: 131  
 Percent Similarity: 59.93% Conservative: 41  
 Best Local Similarity: 45.64% Mismatches: 99  
 Query Match: 39.30% Indels: 16  
 DB: 2 Gaps: 4

US-10-062-624-41 (1-840) x Q9R8A9 (1-278)

QY 1 ATGAATATATAGAAAAATTTCTAGTAAGAACGCGCTTAATCTCATTAATCTCAATCTTACCA 60  
 Db |||||  
 QY 1 MetAsnCysLysLysIleLeuThrAlaLeuIleSerLeuMetTyrSerIlePro 20  
 Db |||||  
 QY 61 TATCAGTCTTTTTCAGATCTCTGAGTTCAGAACTAATATGATAACAAGAGGC----- 114  
 Db |||||  
 QY 21 SerIleSerPheSerAspThrIle-----GlnAspGlyAsnMetGlyGlyAsn 36  
 Db |||||  
 QY 115 TTCTACATTAAGTGCAGAACTCAATCCAGTATATACACACTTTAGAAAATTTCTCTGCTGAA 174  
 Db |||||  
 QY 37 PheTyrIleSerGlyLysTyrValProSerValSerHisPheGlySerPheSerAlaLys 56  
 Db |||||  
 QY 175 GAAACTCTCTAATTAATGGAACAAATTTCTCTCACTAAAAAGTTTTCGGAGTAAAGAAAGAT 234  
 Db |||||  
 QY 57 Glu-----GluSerLysSerThrValGlyValPheGlyLeuLysHisAsp 71  
 Db |||||  
 QY 235 GGTGAT-----ATAACAAAAAGACGATTTTACAAGAGTAGTCCAGGC 279  
 Db |||||  
 QY 72 TrpAspGlySerProIleLeuLysAsnLysHisAlaAspPheThrValProAsnTyrSer 91  
 Db |||||  
 QY 280 ATTGATTTTCAAAATAACTTAATATCAGGATTTTTCAGGAAGTATTGGTTACTCTATGGAC 339  
 Db |||||  
 QY 92 PheArgTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGly 111  
 Db |||||  
 QY 340 GGACCAAGATAGAACTTGAAGTGCATATCAACAATTTAATCCAAAAAACACCGATTAAC 399  
 Db |||||  
 QY 112 GlyProArgIleGluPheGluIleSerTyrGluAlaPheAspValLysSerProAsnIle 131  
 Db |||||  
 QY 400 AATGATCTGATATGAGTGAATATCTATAAATTTTGCATTTATCTCGTAAAGATGCAATG 459  
 Db |||||  
 QY 132 AsnTyrGlnAsnAspAlaHisArgTyrCysAlaLeuSerHisHisThrSerAlaAlaMet 151  
 Db |||||  
 QY 460 GAAGATCAGCAATATGTAGTACTTAAATAATGACGGCATAACTTTTATGTCATTGATGTT 519  
 Db |||||  
 QY 152 GluAlaAspLysPheValPheLeuLysAsnGluGlyLeuIleAspIleSerLeuAlaIle 171  
 Db |||||  
 QY 520 AATACTTCTGATGATATCAGCTGAGGAGTATCTTTTCGTACCATATGCGATGTCAGGT 579  
 Db |||||  
 QY 172 AsnAlaCysTyrAspIleIleAsnAspLysValProValSerProTyrIleCysAlaGly 191  
 Db |||||  
 QY 580 ATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGA 639  
 Db |||||  
 QY 192 IleGlyThrAspLeuIleSerMetPheGluAlaThrSerProLysIleSerTyrGlnGly 211  
 Db |||||  
 QY 640 AAAATAGGTATTAGTTACCTATCACACGAGGTCCTGCGATTTTATTTGTTGGTACTAC 699  
 Db |||||  
 QY 212 LysLeuGlyIleSerTyrSerIleAsnProGluThrSerValPheIleGlyHisPhe 231  
 Db |||||  
 QY 700 CATGCGTATTGGTAAATAATTTGAGAAGATACCTGTAATACTCTCTGATGATTTAAAT 759  
 Db |||||  
 QY 232 HisArgIleIleGlyAsnGluPheArgAspIleProAlaIleValProSerAsnSerThr 251  
 Db |||||  
 QY 760 GATGCTCTCAACACCATCTGTTTCACTACTCTTGAGTCTTGAGTATTTGGGGAGAA 819  
 Db |||||  
 QY 252 ThrIleSerGlyProGlnPheAlaThrValThrLeuAsnValCysHisPheGlyLeuGlu 271  
 Db |||||  
 QY 820 ATTGGAATGAGTTCACCTTC 840  
 Db |||||  
 QY 272 LeuGlyGlyArgPheAsnPhe 278  
 Db |||||

## RESULT 24

Q9R8A8 PRELIMINARY; PRT; 278 AA.  
 ID Q9R8A8  
 AC Q9R8A8;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)



```

Db      72 TrpaspGlySerProIleLeuLysAsnLysHisAlaAspPheThrValProAsnTyrSer 91
Qy      280 ATTGATTTTCAAAATAAATCAAGATTTTCAGGAAGTATTGGTTACTCTATCGAC 339
Db      92 PheArgTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGly 111
Qy      340 GGACCAAGATAGAACTTGAAGTGCATATCAACAATTAATTAATCAAAAACACCGATAC 399
Db      112 GlyProArgIleGluPheGluIleSerTyrGluAlaPheAspValLysSerProAsnIle 131
Qy      400 AATGATCTGATAATGGTGAATCTATCAATCAATTTTGCATTATCTCGTAAGATGCAATG 459
Db      132 AsnTyrGlnAsnAspAlaHisArgTyrCysAlaLeuSerHisThrSerAlaAlaMet 151
Qy      460 GAAGATCAGCATATAGTAGTACTTAAATAATGACGGCATAACTTTTATGCTATTGATGGTT 519
Db      152 GluAlaAspLysPheValPheLeuLysAsnGluGlyLeuIleAspIleSerLeuAlaIle 171
Qy      520 AATAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 579
Db      172 AsnAlaCysTyrAspIleIleAsnAspLysValProValSerProTyrIleCysAlaGly 191
Qy      580 ATAGGACGAGATCTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGA 639
Db      192 IleGlyThrAspLeuIleSerMetPheGluAlaThrSerProLysIleSerTyrGlnGly 211
Qy      640 AAAATAGGTATTAGTATACCTTACACACGAGTCTCTGATTTATTTGGTGGGATAC 699
Db      700 CATGGGTTATTGGTAATAATTTGAGAGATACCTGTGAATCTCTGTTAGTATTAAAT 759
Db      232 HisArgIleIleGlyAsnGluPheArgAspIleProAlaIleValProSerAsnSerThr 251
Qy      760 GATGCTCTCAACACCATCTGCTCAGTACTCTGAGTACTCTGAGTACTCTGAGTACTCT 819
Db      252 ThrIleSerGlyProGlnPheAlaThrValThrLeuAsnValCysHisPheGlyLeuGlu 271
Qy      820 ATTGGAATGAGGTTCCACCTTC 840
Db      272 LeuGlyGlyArgPheAsnPhe 278

```

## RESULT 26

```

ID Q9R8A5 PRELIMINARY; PRT; 278 AA.
AC Q9R8A5;
DC 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FUZZY;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, XJ, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen."
RL Clin. Diagn. Lab. Immunol. 6:392-399 (1999).
DR EMBL: AF082749; AAC64555.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

```

## Alignment Scores:

```

Pred. No.: 5.8e-40 Length: 278
Score: 582.00 Matches: 131

```

```

Percent Similarity: 59.93% Conservative: 41
Best Local Similarity: 45.64% Mismatches: 99
Query Match: 39.30% Indels: 16
DB: 2 Gaps: 4

```

US-10-062-624-41 (1-840) x Q9R8A5 (1-278)

```

Qy      1 ATGAATTATAAGAAAATTTCTAGTAAGAGCGCGTAAATCTCATTAATGTCTAATCTTACCA 60
Db      1 MetAsnCysLysIleLeuIleThrAlaLeuIleSerLeuMetTyrSerIlePro 20
Qy      61 TATCAGTCTTTTTCAGATCCCTGAGGTTCAGAACTAATGATATAACAAGAAGGC----- 114
Db      21 SerIleSerPheSerAspThrIle-----GlnAspGlyAsnMetGlyGlyAsn 36
Qy      115 TTCTACATTAAGTCAAAAGTACAATCCAAAGTATATACACATTTTACAAAAATTTCTCTGCTGAA 174
Db      37 PheTyrIleSerGlyLysTyrValProSerValSerHisPheGlySerPheSerAlaLys 56
Qy      175 GAAACTCTTATTAATGAACAATAATCTCTCACTAAAAAAGTTTTCGGAATAAAGAAAGAT 234
Db      57 Glu-----GluSerLysSerThrValGlyValPheGlyLeuLysHisAsp 71
Qy      235 GGTGAT-----ATACAAAAAAGACGATTTTACAGAGTAGTCCAGGC 279
Db      72 TrpAspGlySerProIleLeuLysAsnLysHisAlaAspPheThrValProAsnTyrSer 91
Qy      280 ATTGATTTTCAAAATAAATCAAGATTTTTCAGGAAGTATTGGTTACTCTATCGAC 339
Db      92 PheArgTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGly 111
Qy      340 GGACCAAGATAGAACTTGAAGTGCATATCAACAATTTTAAATTTTCAAAAACACCGATAC 399
Db      112 GlyProArgIleGluPheGluIleSerTyrGluAlaPheAspValLysSerProAsnIle 131
Qy      400 AATGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 459
Db      132 AsnTyrGlnAsnAspAlaHisArgTyrCysAlaLeuSerHisThrSerAlaAlaMet 151
Qy      460 GAAGATCAGCAATATAGTAGTACTTAAATAATGACGGCATAACTTTTATGCTATTGATGGTT 519
Db      152 GluAlaAspLysPheValPheLeuLysAsnGluGlyLeuIleAspIleSerLeuAlaIle 171
Qy      520 AATAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 579
Db      172 AsnAlaCysTyrAspIleIleAsnAspLysValProValSerProTyrIleCysAlaGly 191
Qy      580 ATAGGACGAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGA 639
Db      192 IleGlyThrAspLeuIleSerMetPheGluAlaThrSerProLysIleSerTyrGlnGly 211
Qy      640 AAAATAGGTATTAGTATACCTTACACACGAGTCTCTGATTTATTTGGTGGGATAC 699
Db      212 LysLeuGlyIleSerTyrSerIleAsnProGluThrSerValPheIleGlyGlyHisPhe 231
Qy      700 CATGGGCTTATGCTAATAAATTTGAGAGATACCTCTGTAATACCTCTGATGATTAAT 759
Db      232 HisArgIleIleGlyAsnGluPheArgAspIleProAlaIleValProSerAsnSerThr 251
Qy      760 GATGCTCTCAACACCATCTGCTCAGTACTCTGAGTACTCTGAGTACTCTGAGTACTCT 819
Db      252 ThrIleSerGlyProGlnPheAlaThrValThrLeuAsnValCysHisPheGlyLeuGlu 271
Qy      820 ATTGGAATGAGGTTCCACCTTC 840
Db      272 LeuGlyGlyArgPheAsnPhe 278

```

## RESULT 27

```

Q9R8A6 PRELIMINARY; PRT; 278 AA.
AC Q9R8A6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

```



```

Db 72 TrpAspGlySerProIleLeuLysSerHisAlaAspPheThrValProAsnTyrSer 91
Qy 280 ATTGATTTCCTCAAAATTAATATATAGAGATTTTCAGGAAGTATTGGTTACTTATGAC 339
Db 92 PheArgTyrGluAsnAspPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGly 111
Qy 340 GGACCAAGATAGACTTGAAGTCGCATATCAACATTTAATCCAAAACACCGATAC 399
Db 112 GlyProArgIleGluPheGlySerTyrGluAlaPheAspValLysSerProAsnIle 131
Qy 400 AATGATACCTGATAATGGTGAATACATTAACATTTTGCATTATCTCGTAAAGATCAATG 459
Db 132 AsnTyrGlnAsnAspAlaHisArgTyrCysAlaLeuSerHisHisThrSerAlaAlaMet 151
Qy 460 GAAGATCAGCAATATGTAGTACTTAAATGACGGCATAACTTTTATGCTATTGATGTT 519
Db 152 GluAlaAspLysPheValPheLeuLysAsnGluGlyLeuIleAspIleSerLeuAlaIle 171
Qy 520 AATACTTCTGATGACATTTACAGCTGACGAGATATCTTCGTACCATATGCGATGCGAGT 579
Db 172 AsnAlaCysTyrAspIleAsnAspLysValProValSerProTyrIleCysAlaGly 191
Qy 580 ATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAGGA 639
Db 192 IleGlyThrAspLeuIleSerMetPheGluAlaThrSerProLysIleSerTyrGlnGly 211
Qy 640 AAAATAGTATTAGTATACCTATCACACGAGAGTCTCTGCTATTATTGGTGGATACATAC 699
Db 212 LysLeuGlyIleSerTyrSerIleAsnProGluThrSerValPheIleGlyHisPhe 231
Qy 700 CATGGCGTTTGGTAATAATTTGAGAAGATACCTGTAAATCTCTGCTGTAGTATTAAAT 759
Db 232 HisArgIleGlyAsnGluPheArgAspIleProAlaIleValProSerAsnSerThr 251
Qy 760 GATGCTCCTCAACACCATCTCTCACTAACTCTTGAGTGGATACCTTTGGCGGAGAA 819
Db 252 ThrIleSerGlyProGlnPheAlaThrValThrLeuAsnValCysHisPheGlyLeuGlu 271
Qy 820 ATTGGAATGAGGTTCCACCTTC 840
Db 272 LeuGlyGlyArgPheAsnPhe 278

RESULT 29
O52106 PRELIMINARY; PRT; 278 AA.
AC O52106;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE OMP-1E.
GN OMP-1E.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_taxid=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikhisia Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
RL are encoded by a polymorphic multigene family.";
RN Infect. Immun. 66:132-139(1998).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RA Allemen A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
RL the tribe Ehrlichiae.";
RN Biochem. Biophys. Res. Commun. 247:636-643(1998).
DB EMBL; U7291; AAC02939.1; -

```

```

DR EMBL; AF062761; AAC26719.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 278 AA; 30543 MW; E321B3CA259B87FD CRC64;

Alignment Scores:
Pred. No.: 1.5e-39 Length: 278
Score: 577.00 Matches: 126
Percent Similarity: 60.55% Conservative: 49
Best Local Similarity: 43.60% Mismatches: 94
Query Match: 38.96% Indels: 20
Gaps: 7

US-10-062-624-41 (1-840) x O52106 (1-278)
Qy 1 ATGAATTATATAAGAAATTTCTAGTAAGAACGCGTTAATCTCATTAAATGCTCAATCTTACCA 60
Db 1 MetAsnCysLysLysPhePheIleThrThrAlaLeuValSerLeuMetSerPheLeuPro 20
Qy 61 TATCAGCTTTTTCAGATCTCTGTAGTTTCAAGAACTAATGATAACAAGAAAGGC---TTC 117
Db 21 GlyIleSerPheSerAspProVal-----GlnGlyAspAsnIleSerGlyAsnPhe 37
Qy 118 TACATTAGTCAAAAGTCAATCCAACTATATCACACTTTAGAAAATTTCTCTGCTGAAGAA 177
Db 38 TyrValSerGlyLysTyrMetProSerAlaSerHisPheGlyMetPheSerAlaLysGlu 57
Qy 178 ACTCCTTAAATGAACAATAATCTCTCACTAAAAAGTTTTTCGGACTAAAGAAAGAT--- 234
Db 58 -----GluLysAsnProThrValAlaLeuTyrGlyLeuLysGlnAspTrp 72
Qy 235 -----GGTGTATATAACAAAAAGACGATTTTACAGAGTAGTACCTCCAGGCATT 282
Db 73 GluGlyIleSerSerSerHisAsnAspAsnHisPheAsnAsnLysGlyTyrSerPhe 92
Qy 283 GATTTTCAATAACTTAATATCAGGATTTTCAGGAAGTATTGGTTACTTATGACCGGA 342
Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGlyGly 112
Qy 343 CCAAGAATAGAACTTGAAGCTGCATATCAACAAATTTAATCCAAAAACACCGATAACAAT 402
Db 113 ProArgValGluPheGluValSerTyrGluThrPheAspValLysAsnGlnGlyAsnAsn 132
Qy 403 GATACTGATATGTGTAATCTATAAACATTTTGCATTATCTCGTAAAGAT----- 453
Db 133 TyrLysAsnAspAla-----HisArgTyrCysAlaLeuGlyGlnGlnAsnSerGly 150
Qy 454 GCAATGGAAGATCAGCAATATGATAGTACTTAAATGACGCGCATAACTTTTATGTCATTG 513
Db 151 IleProLysThrSerLysTyrValLeuLeuLysSerGluGlyLeuLeuAspIleSerPhe 170
Qy 514 ATGGTTAATCTCTGATGATACATTACAGCTGACGAGGAGTATCTTTCGTACCATATGCGATG 573
Db 171 MetLeuAsnAlaCysTyrAspIleIleAsnGlnSerIleProLeuSerProTyrIleCys 190
Qy 574 GCAGGTATAGGAGCAGATCTTATCACTATTTTAAAGACCTCAATCTAAAATTTGCTTAC 633
Db 191 AlaGlyValGlyThrAspLeuIleSerMetPheGluAlaThrAsnProLysIleSerTyr 210
Qy 634 CAAGGAAAAATAGTATTAGTTACCTATATCACACAGAGTCTCTGCAATTTATTGGTGA 693
Db 211 GlnGlyLysLeuGlyLeuSerTyrSerIleAsnProGluAlaSerValPheIleGlyGly 230
Qy 694 TACTACCATGCGCTTATTGGTAAATAATTTGAGAAGATACCTGTAAATACCTCTCTGTAGTA 753
Db 231 HisPheHisLysValIleGlyAsnGluPheArgAspIleProThrLeuLysAlaPheVal 250
Qy 754 TTAATGATGCTCCTCAAAACACATCTGCTTACGTAATCTTTGAGTTGGATCTTTGGC 813
Db 251 ThrSerAlaThrProAspLeu--AlaIleValThrLeuSerValCysHisPheGly 269
Qy *814 GGAGAAATGGAATGAGGTTCCACCTTC 840

```

```

Db      270 IleGluLeuGlyGlyArgPheAsnPhe 278
RESULT 30
Q9F472
ID      Q9F472      PRELIMINARY;      PRT;      278 AA.
AC      Q9F472;
DT      01-MAR-2001 (TrEMBLrel. 16, Created)
DT      01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT      01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE      P28-7.
GN      P28-7.
OS      Ehrlichia canis.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC      Anaplasmataceae; Ehrlichia.
OX      NCBI_TaxID=944;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=JAKE;
RX      MEDLINE=99242757; PubMed=10225842;
RA      McBride J.W., Yu, X.J., Walker D.H.;
RT      "Molecular cloning of the gene for a conserved major immunoreactive
RT      28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT      antigen.";
RT      Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RL      [2]
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=JAKE;
RX      MEDLINE=20432107; PubMed=10974556;
RA      McBride J.W., Yu X.J., Walker D.H.;
RT      "A conserved, transcriptionally active p28 multigene locus of
RT      Ehrlichia canis.";
RL      Gene 254:245-252(2000).
DR      EMBL; AF082744; AAC64550.2; -.
DR      InterPro; IPR002566; Surface_Ag_msp4.
DR      Pfam; PF01617; Surface_Ag_2; 1.
SQ      SEQUENCE 278 AA; 30455 MW; 2411CAAB4C56CA74 CRC64;

Alignment Scores:
Pred. No.:      1.5e-39      Length:      278
Score:      577.00      Matches:      130
Percent Similarity:      59.58%      Conservative:      41
Best Local Similarity:      45.30%      Mismatches:      100
Query Match:      38.96%      Indels:      16
DB:      2      Gaps:      4

US-10-062-624-41 (1-840) x Q9F472 (1-278)

QY      1 ATGAATTATAAGAAATCTAGTAAGACGGGTTAATCTCTATTATGTCATCTTACCA 60
Db      1 MetAsnCysLysIleLeuLeuThrAlaLeuLeuSerLeuMetTyrSerIlePro 20
QY      61 TATCAGTCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGTAACAAGAAGGC----- 114
Db      21 SerIleSerPheSerAspThrIle-----GlnAspGlyAsnMetGlyGlyAsn 36
QY      115 TTCTACATTAGTCGAAGTACAACTCAAGTATATCACATTTAGAAAATCTCTGCTGAA 174
Db      37 PheTyrIleSerGlyLysTyrValProSerValSerHisPheGlySerPheSerAlaTys 56
QY      175 GAAACTCTTATTAAATGGAACAATCTCTCACTTAAAAAGTTTCGGACTTAAAGAAAGAT 234
Db      57 Glu-----GluSerLysSerThrValGlyValPheGlyLeuLysHisAsp 71
QY      235 GTGTAT-----ATAACAAAAAGACGATTTTCAAGAGTAGTCTCAGGC 279
Db      72 TrpAspGlySerProIleLeuLysAsnLysHisAlaAspPheAlaValProAsnTyrSer 91
QY      280 ATTGATTTCAAAATTAATATATCAGGATTTTCAGGAAGTATGTTACTCTATGGAC 339
Db      92 PheArgTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGly 111
QY      340 GGACCAAGATAGAACTTGAAGTCGATATCAACAATTTAAATCCAAAAAACCCGATAAC 399

```

```

Db      112 GlyProArgIleGluPheGluLeuSerTyrGluAlaPheAspValLysSerProAsnIle 131
QY      400 AATGATCTACTGATAATGGTGAATACTATATAACATTTTGCATTTATCTCGTAAGATGCAATG 459
Db      132 AsnTyrGlnAsnAspAlaHisArgTyrCysAlaLeuSerHisHisThrSerAlaAlaMet 151
QY      460 GAAGATCAGCAATATGTAGTACTTTAAAAATGACGCGCATAACTTTTATGCTCATGTGTT 519
Db      152 GluAlaAspLysPheValPheLeuLysAsnGluGlyLeuIleAspIleSerLeuAlaIle 171
QY      520 AATACTTCTATGACATTACAGCTGAAGGAGTATCTTTTCGTACCATATGTCAGGT 579
Db      172 AsnAlaCysTyrAspIleIleAsnAspLysValProValSerProTyrIleCysAlaGly 191
QY      580 ATAGAGCAGATCTTATCATTCTTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGA 639
Db      192 IleGlyThrAspLeuIleSerMetPheGluAlaThrSerProLysIleSerTyrGlnGly 211
QY      640 AAAATAGGTATTAGTTACCTTATCACACCAAGTCTCTGCATTTATTGGTGGATAC 699
Db      212 LysLeuGlyIleSerTyrSerIleAsnProGluThrSerValPheIleGlyHisPhe 231
QY      700 CATGCGCTTATTGGTAAATAATTTGAGAAGATACCTGTAACTCTCTGTAGTATAAT 759
Db      232 HisArgIleIleGlyAsnGluPheArgAspIleProAlaIleValProSerAsnSerThr 251
QY      760 GATGCTCTCTAAACACACATCTGTTTCAGTAACTCTTGCATCTTGGATATCTTGGCGGAGAA 819
Db      252 ThrIleSerGlyProGlnPheAlaThrValThrLeuAsnValCysHisPheGlyLeuGlu 271
QY      820 ATTGGAATGAGGTTCACCTTC 840
Db      272 LeuGlyGlyArgPheAsnPhe 278

RESULT 31
Q9S6H1
ID      Q9S6H1      PRELIMINARY;      PRT;      281 AA.
AC      Q9S6H1;
DT      01-MAY-2000 (TrEMBLrel. 13, Created)
DT      01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT      01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE      Major antigenic protein 1.
GN      MAP1.
OS      Cowdria ruminantium.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC      Anaplasmataceae; Ehrlichia.
OX      NCBI_TaxID=779;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=LEMCOT3;
RX      MEDLINE=99216274; PubMed=10198207;
RA      Sulsona C.R., Mahan S.M., Barbet A.F.;
RT      "The map1 Gene of Cowdria ruminantium Is a Member of a Multigene
RT      Family Containing Both Conserved and Variable Genes.";
RL      Biochem. Biophys. Res. Commun. 257:300-305(1999).
DR      EMBL; AF125277; AAD26350.1; -.
DR      InterPro; IPR002566; Surface_Ag_msp4.
DR      Pfam; PF01617; Surface_Ag_2; 1.
SQ      SEQUENCE 281 AA; 30315 MW; 9A33A1C264A438E1 CRC64;

Alignment Scores:
Pred. No.:      9.14e-39      Length:      281
Score:      567.50      Matches:      134
Percent Similarity:      57.58%      Conservative:      37
Best Local Similarity:      45.12%      Mismatches:      93
Query Match:      38.32%      Indels:      33
DB:      2      Gaps:      8

US-10-062-624-41 (1-840) x Q9S6H1 (1-281)

QY      1 ATGAATTATAAGAAATTTCTAGTAAGACGGGTTAATCTCTATTATGTCATCTTACCA 60
Db      1 MetAsnCysLysIlePheIleThrSerThrLeuIleSerLeuValSerPheLeuPro 20

```



```

QY 61 TATCAGTCTTTGCGAGT-----CCTGTAGGTTCAAGAACTAAT 99
Db 21 GlyValSerPheSerAspAlaIleGlnGluAspSerProValGlySer----- 37
QY 100 GATACAAAGAGGCTTCTACATTAGTCAAGTACAATCCAAAGTATATACACTTTAGA 159
Db 38 -----ValTyrIleSerAlaLysTyrMetProThrAlaSerHisPheGly 52
QY 160 AAATCTCTGCTGAAGAACTCTTAAATTAATGAACAAATTTCTCACAATAAAAGTTTTC 219
Db 53 LysMetSerIleLysGluAspSerArgAsp-----ThrLysValValPhe 67
QY 220 GACTAAAGAAGATGGTGATATATACAAAAAAGACGAT-----TTTACAAGA 267
Db 68 GlyLeuLysLysAspTrpAspGlyValLysThrSerSerSerAsnThrIlePheThrGlu 87
QY 268 GTAGCTCCAGGCAATGATTTTCAAAATACITTAATATCAGGATTTTCAGGAGTATTCGT 327
Db 88 LysAspTyrSerPheLysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGly 107
QY 328 TACTCTATGGCGGACCAAGATGAAGTCTGAAGCTGCATATCAACAATTTAATCCAAA 387
Db 108 TyrSerMetAsnGlyProArgIleGluPheGluIleSerTyrGluThrPheAspValLys 127
QY 388 AACCCGATACAAATGATGATCTGATTAATGGTGAATACTAT-----AAACATTTTGCATTA 441
Db 128 AsnProGlyGlyAsnTyrLysAsnAspAlaHisMetTyrCysAlaLeuAspThrAlaThr 147
QY 442 TCTCTGAAGATGCAATGGAGATCAGCAATATGTAGTACTTAAATAAGCGGCATACT 501
Db 148 SerSerGlyGlyAlaAlaAlaSerThrSer---ValMetValLysAsnGluAsnLeuThr 166
QY 502 TTTATGTCTATGATGGTAAATCTTGTATGATGATACATTTACAGCTGAAGGAGTATCTTCGTA 561
Db 167 AspileSerLeuMetLeuAsnAlaCysTyrAspileMetLeuAspGlyMetProValSer 186
QY 562 CCAATATGATGTGAGGTATAGGACAGATCTTATCACTATTTTAAAGACCTCAATCTA 621
Db 187 ProTyrValCysAlaGlyIleGlyThrAspLeuValSerValIleAsnSerThrAsnPro 206
QY 622 AAATTTGCTTACCAGGAAATAGGTATTTAGTACCTTATCACCATCACACGAGAGTCTCTCCA 681
Db 207 LysLeuSerTyrGlnGlyLysLeuGlyIleSerTyrSerIleAsnProGluAlaSerIle 226
QY 682 TTTATGTGTGATCTACATCGCGCTTATGGTAAATAATTTGAGAAGATACCTCTAATA 741
Db 227 PheIleGlyGlyHisPheHisArgValIleGlyAsnGluPheLysAspile-----Thr 244
QY 742 ACTCTGTAGTATTAATGATGCTCTCTCAAAACCACATCT-----GCTTCAGTA 789
Db 245 ThrSerLysIlePheAsnThrSerAsnThrGlyGlyAlaThrProGlyPheAlaSerAla 264
QY 790 ACTCTTGAGTTGGATCTTTGGCGGAGAAATTTGAAGTTCACCTTC 840
Db 265 ThrLeuAspValCysHisPheGlyIleGluIleGlyIleGlyIleGlyIleGlyIleGlyIle 281
RESULT 32
O69197 PRELIMINARY; PRT; 271 AA.
AC O69197
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P28-20 precursor (Major outer membrane protein P28-1).
GN P28-20 OR P28-1.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;

```

```

RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RA Yu X.J., McBride J.W., Zhang X., Walker D.H.;
RT "Characterization of the complete transcriptionally active Ehrlichia
chaffeensis 28 kDa outer membrane protein multigene family.";
RL Gene 248:29-68(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
Membrane Protein Multigene Family in Ehrlichia canis and E.
chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF068234; AAC19134.2; -.
DR EMBL; U72291; AAK28674.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 271 P28-20.
SQ SEQUENCE 271 AA; 30050 MW; 84EDF65DF3DB61C CRC64;

Alignment Scores:
Pred. No.: 1,16-38 Length: 271
Score: 566.50 Matches: 123
Percent Similarity: 61.35% Conservative: 50
Best Local Similarity: 43.62% Mismatches: 94
Query Match: 38.25% Indels: 15
DB: 2 Gaps: 9

US-10-062-624-41 (1-840) x O69197 (1-271)
QY 1 ATGAATTATAGAATAATTTCTAGTAAGAAGCGGTTAATCTCATTAATGCTCAATCTTACCA 60
Db 1 MetAsnTyrLysLysPheValValGlyValAlaLeuAlaThrLeuLeuSerPheLeuPro 20
QY 61 TATCAGTCTTTTGGAGATCTCTGTAGTTTCAAGAACTAATGATAACAAGAAGGCTTCTAC 120
Db 21 AspAsnSerPheSerAsp-----AlaAsnValProGluGlyArgLysGlyPheTyr 37
QY 121 ATTAGTCAAGTACATCCAGTATATCATCTTTAGAAAATTTCTCTGCTGAGAAACT 180
Db 38 ValGlyThrGlnTyrLysValGlyValProAsnPheSerAsnPheSerAlaGluGluThr 57
QY 181 CCTATTATGAACAAATTTCTCTCATAAATAAATTTTCGACCTAAAGAAAGAT---GGT 237
Db 58 -----LeuProGlyLeuThrLysSerIlePheAlaLeuGlyLeuAspLysSer 73
QY 238 GATATAACAAAAAGACGATTTTACAAGAGTACTCCAGGCATTTGATTTTCAAAATAAC 297
Db 74 SerIleSerAspHisAlaGlyPheThrGln---AlaTyrAsnProThrTyrAlaSerAsn 92
QY 298 TTAATATCAGGATTTTTCAGGAAGTATTGGTTACTCTATGACCGGACCAAGAAATGAACTT 357
Db 93 ---PheAlaGlyPheGlyGlyValIleGlyTyrTyrValAsnAspPheArgValGluPhe 111
QY 358 GAACCTCATATCAACAATTTAATCCAAAAACACCCGATCAATCAATGATGATAATGCT 417
Db 112 GluGlyAlaTyrGluAsnPheGluProGluArg-----GlnTyrTyrProGluGlyGly 129
QY 418 GAATACTATAACATTTTGTGATTTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTA 477
Db 130 GluSerHisLysPhePheAlaLeuSerArgGluSerThrValGlnAspAsnLysPheIle 149
QY 478 GTACTTAAATGACGCAATACTTTTATGTCATTGATGGTTAATATCTTGTATGACATT 537
Db 478 GTACTTAAATGACGCAATACTTTTATGTCATTGATGGTTAATATCTTGTATGACATT

```





Db 93 LysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGlyTyrSerMetGlyGly 112  
 QY 343 CAAGAATAAGACTGTAATCAACAAATTAATCCAAAACCCGATCAACAT 402  
 Db 113 ProArgIleGluPheGluValSerTyrGluThrPheAspValLeuAsnGlnGlyGlyAsn 132  
 QY 403 GATACCTGATAAGTGGTAACTATATAACATTTTGGCAATTCCTGTAAGCATCAATGGAA 462  
 Db 133 TyrLysAsnAspAla-----HisArgTyrCysAlaLeuAspArgLysAlaSerSerThr 150  
 QY 463 GAT-----CAGCAATATAGTACTATAAATGACGGCATCACTTTATGTCATTG 513  
 Db 151 AsnAlaThrAlaSerHisTyrValLeuLeuLysAsnGluGlyLeuAspIleSerLeu 170  
 QY 514 ATGGTTAATACCTTGTATGACATTAACAGTGAAGAGTATCTTCTGATCATATGTCATG 573  
 Db 171 MetLeuAsnAlaCysTyrAspValSerGluGlyIleProPheSerProTyrIleCys 190  
 QY 574 CGAGGTATAGGACGATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTAC 633  
 Db 191 AlaGlyValGlyThrAspLeuIleSerMetPheGluAlaIleAsnProLysIleSerTyr 210  
 QY 634 CAAGAAAATAGTATTAGTACCTATCACACCAAGATCTCTGCAATTTATGTCGA 693  
 Db 211 GlnGlyLysLeuGlyLeuSerTyrSerIleAsnProGluAlaSerValPheValGlyGly 230  
 QY 694 TACTACCTGGCGTTATTGGTAAATAATTGAGAGATACCTGTTAATA----- 741  
 Db 231 HisPheHisLysValAlaGlyAsnGluPheArgAspIleSerThrLeuLysAlaPheAla 250  
 QY 742 ACTCTGTAGTATTAATGATCTCTCAACACATCTGCTTCAGTAACTCTTGCAGTT 801  
 Db 251 ThrPro-----SerSerAlaAlaThrProAspLeuAlaThrValThrLeuSerVal 267  
 QY 802 GGATACCTTTGGCGGAAATTTGGAATGAGTTCAGGTTCCACCTTC 840  
 Db 268 CysHisPheGlyValGluLeuGlyGlyArgPheAsnPhe 280  
 RESULT 36  
 Q46328 PRELIMINARY; PRT; 281 AA.  
 AC Q46328;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE Major antigenic protein 1.  
 GN NAPI.  
 OS Cowdria ruminantium.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Anaplasmataceae; Ehrlichia.  
 OX NCBI\_TaxID=779;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ANTIGUA;  
 RX MEDLINE=96400830; PubMed=8807206;  
 RA Reddy R.G., Sulsona C.R., Harrison R.H., Mahan S.M., Burrige M.J.,  
 RA Barbet A.F.;  
 RT "Sequence heterogeneity of the major antigenic protein 1 genes from  
 RT Cowdria ruminantium isolates from different geographical areas.";  
 RL Clin. Diagn. Lab. Immunol. 3:417-422(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ANTIGUA;  
 RX MEDLINE=99216274; PubMed=10198207;  
 RA Sulsona C.R., Mahan S.M., Barbet A.F.;  
 RT "The mapi gene of Cowdria ruminantium is a member of a multigene  
 RT family containing both conserved and variable genes.";  
 RL Biochem. Biophys. Res. Commun. 257:300-305(1999).  
 RL EMBL; U50830; AAC4414.1; -;  
 DR EMBL; AF125279; AAD26354.1; -;  
 DR InterPro; IPR002566; Surface Ag\_msp4.  
 DR Pfam; PF01617; Surface Ag 2; 1.  
 SQ SEQUENCE 281 AA; 30355 MW; 8FE99C5964A43D6A CRC64;

Alignment Scores: 2.86e-38 Length: 281  
 Pred. No.: 561.50 Matches: 133  
 Score: 57.24% Conservative: 37  
 Percent Similarity: 44.78% Mismatches: 94  
 Best Local Similarity: 37.91% Indels: 33  
 Query Match: 2 Gaps: 8  
 DB:

US-10-062-624-41 (1-840) x Q46328 (1-281)

QY 1 ATGAATATAAGAAATCTAGTAAGACGCGTTAAATCTCATTAATCTCAATCTTACCA 60  
 Db 1 MetAsnCysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 20  
 QY 61 TATCAGTCTTTTGCAGAT-----CCTGTAGGTTCAAGAACTAAT 99  
 Db 21 GlyValSerPheSerAspValIleGlnGluAspSerSerProValGlySer----- 37  
 QY 100 GATAACAAGAGGCTTCTACATTAGTCAAGTACAAATCCAAAGTATATACACTTTAGA 159  
 Db 38 -----ValTyrIleSerAlaLysTyrMetProThrAlaSerHisPheGly 52  
 QY 160 AAATTTCTCTGCTGAAGAACTCCTATTATTAAGAACAAATCTCTCACTAAAGAAAGTTTC 219  
 Db 53 LysMetSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 67  
 QY 220 GGACTAAAGAAAGATGGTGATATACAAACAAAAGACGAT-----TTTCAAGA 267  
 Db 68 GlyLeuLysLysAspTyrAspGlyValLysThrSerSerSerSerSerSerSerSerSer 87  
 QY 268 GTAGCTCCAGCATTTGATTTTCAAAATTAATTAATATCAGGATTTTCAGGAGTATTCGT 327  
 Db 88 LysAspTyrSerPheLysTyrGluAsnAsnProPheLeuGlyPheAlaGlyAlaIleGly 107  
 QY 328 TACTCTTAGGACGACCAAGATAGAACTTGAAGTGCATATCAACAATTTAATCCAAA 387  
 Db 108 TyrSerMetAsnGlyProArgIleGluPheGluLysSerTyrGluThrPheAspValLys 127  
 QY 388 AACCCGATAACATGATGATGATTAATGGTGAATACTAT-----AACATTTTGCATTA 441  
 Db 128 AsnProGlyGlyAsnTyrLysAsnAspAlaHisMetTyrCysAlaLeuAspThrAlaThr 147  
 QY 442 TCTCTGAAGATGCAATGGAAGATCAGCAATATGATAGTACTTAAAGATCAGGCATAACT 501  
 Db 148 SerSerGlyGlyAlaAlaAlaSerThrSer-----ValMetValLysAsnGluAsnLeuThr 166  
 QY 502 TTTATGTCATTTGATGTTTAAATCTTGTATGATGATGATGATGATGATGATGATGATGAT 561  
 Db 167 AspIleSerLeuMetLeuAsnAlaCysTyrAspIleMetLeuAspGlyMetProValSer 186  
 QY 562 CCATATGATGTCAGGATATAGGACGATCTTATCACTATTTTAAAGACCTCAATCTA 621  
 Db 187 ProTyrValCysAlaGlyIleGlyThrAspLeuValSerValIleAsnSerThrAsnPro 206  
 QY 622 AAATTTGCTTACCAGGAAATAGGATATAGTATACCTATCACCCTATCACCAGCACTCTGCA 681  
 Db 207 LysLeuSerTyrGlnGlyLysLeuGlyLysSerTyrSerIleAsnProGluAlaSerIle 226  
 QY 682 TTTATGTTGGATACCTACCATGGCGCTTATTTGTTAAATAAATTTGAGAAGATCCTCTAATA 741  
 Db 227 PheIleGlyGlyHisPheHisArgValIleGlyAsnGluPheLysAspIle-----Thr 244  
 QY 742 ACTCTGTAGTATTAATGATGCTCCTCAACACCATCT-----GCTTCAGTA 789  
 Db 245 ThrSerLysIlePheAsnThrSerAsnThrGlyGlyAlaThrProGlyPheAlaSerAla 264  
 QY 790 ACTCTGAGCTGGTATCTTTGGCGGAGAAATTTGGAATGGAATGGAATGGAATGGAATGGAAT 840  
 Db 265 IleLeuAspValCysHisPheGlyIleGluIleGlyGlyArgPheValPhe 281  
 RESULT 37  
 Q46333





US-10-062-624-41 (1-840) x 09AFA1 (1-284)

Search completed: July 8, 2003, 10:48:29  
Job time: 102 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: July 4, 2003, 00:10:02 ; Search time 1655 Seconds  
(without alignments)  
14771.213 Million cell updates/sec  
Title: US-10-062-624-41  
Perfect score: 840  
Sequence: 1 atgaattataagaataattct.....ttggaatgaggttcacattc 840  
Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.hcg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sv.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.ey.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	840	100.0	6913	1	AF324792
2	840	100.0	11329	1	AF082742 Ehrlichia
3	840	100.0	28254	1	AF078553 Ehrlichia
4	600.2	71.5	27190	1	ECU72291 Ehrlichia C
5	528.2	62.9	3507	1	AF125276 Cowdria r
6	528.2	62.9	3535	1	AF125274 Cowdria r
7	528.2	62.9	3538	1	AF125277 Cowdria r
8	528.2	62.9	3541	1	AF125279 Cowdria r
9	528.2	62.9	3551	1	AF125275 Cowdria r
10	528.2	62.9	3572	1	AF125278 Cowdria r
11	457.6	54.5	2362	1	AF125278 Cowdria r
12	238.4	28.4	834	1	AF082745 Ehrlichia
13	238.4	28.4	834	1	AF082746 Ehrlichia
14	238.4	28.4	834	1	AF082747 Ehrlichia
15	238.4	28.4	834	1	AF082748 Ehrlichia
16	238.4	28.4	834	1	AF082749 Ehrlichia
17	238.4	28.4	834	1	AF082750 Ehrlichia
18	234	27.9	1342	1	AF393388 Ehrlichia
19	233.2	27.8	1315	1	AF393392 Ehrlichia
20	228.6	27.2	830	6	AX042315 Sequence
21	228.6	27.2	4683	1	AF062761 Ehrlichia
22	228.2	27.2	1280	1	AF393389 Ehrlichia
23	222.6	26.5	863	1	AF355200 Cowdria r
24	218.4	26.0	1243	1	AF077733 Ehrlichia
25	218.4	26.0	1283	1	AF077732 Ehrlichia
26	218.4	26.0	1307	1	AF393394 Ehrlichia
27	218.4	26.0	1309	1	AF393390 Ehrlichia
28	215.2	25.6	1189	1	AF393391 Ehrlichia
29	213.6	25.4	1282	1	AF393395 Ehrlichia
30	213.6	25.4	1312	1	AF077734 Ehrlichia
31	212.8	25.3	861	6	AX042312 Sequence
32	212.4	25.3	1467	1	CRMAP1 X74250 C.ruminanti
33	212	25.2	1286	1	AF393393 Ehrlichia
34	207.2	24.7	843	6	AX042314 Sequence
35	206.6	24.6	1263	1	CRUS0830 US0830 Cowdria rum
36	205	24.4	837	6	AX042313 Sequence
37	203.8	24.3	842	6	AX042307 Sequence
38	201.2	24.0	1564	1	CRU49843 U49843 Cowdria rum
39	199.4	23.7	1263	1	CRUS0831 US0831 Cowdria rum
40	196.8	23.4	873	1	AY028378 Cowdria r
41	195.4	23.3	980	1	AF077735 Ehrlichia
42	195	23.2	864	6	AX042305 Sequence
43	195	23.2	1265	1	CRUS0833 US0833 Cowdria rum
44	194.8	23.2	1278	1	CRUS0832 US0832 Cowdria rum
45	194.4	23.1	816	1	AF368013 Cowdria r

ALIGNMENTS

RESULT 1  
AF324792  
LOCUS  
DEFINITION  
AF324792 6913 bp DNA linear BCT 11-APR-2001  
Ehrlichia canis phosphoribosylaminoimidazole carboxylase (purK)  
gene, complete cds; major outer membrane protein gene cluster 2,  
complete sequence; and u6 gene, partial cds.  
ACCESSION  
AF324792  
VERSION  
AF324792.1 GI:13591681  
KEYWORDS  
Ehrlichia canis.  
ORGANISM  
Ehrlichia canis  
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
Anaplasmataceae; Ehrlichia.  
REFERENCE  
1 (bases 1 to 6913)  
AUTHORS  
Ohashi,N., Unver,A., Zhi,N. and Rikihisa,Y.



Qy 481 CTTAAATATGCGGATACCTTTTATGTATGATGTTAATATCTGCTATGACATTACA 540  
Db |||||  
2236 CTTAAATATGCGGATACCTTTTATGTATGATGTTAATATCTGCTATGACATTACA 2295  
Qy 541 GCTGAAGGATCTTTTGTACCATATGTCATGTCACGATATAGGAGCAGATCTTATCACT 600  
Db |||||  
2296 GCTGAAGGATCTTTTGTACCATATGTCATGTCACGATATAGGAGCAGATCTTATCACT 2355  
Qy 601 ATTTTAAAGACCTCAATCTAAATTTTGTCTTACCAAGGAAAAATAGGTATTAGTTACCT 660  
Db |||||  
2356 ATTTTAAAGACCTCAATCTAAATTTTGTCTTACCAAGGAAAAATAGGTATTAGTTACCT 2415  
Qy 661 ATCACCAGAGTCTCTGCATTTATGCTGGATATACCATGCGGTTATGGTAATAA 720  
Db |||||  
2416 ATCACCAGAGTCTCTGCATTTATGCTGGATATACCATGCGGTTATGGTAATAA 2475  
Qy 721 TTTGAGAAGATACCTGTAATTAACCTCTGCTAGTATTAAATGATGCTCTCAACACCATCT 780  
Db |||||  
2476 TTTGAGAAGATACCTGTAATTAACCTCTGCTAGTATTAAATGATGCTCTCAACACCATCT 2535  
Qy 781 GCTTCAGTAACCTCTGAGCTTGGATACCTTTGGCGGAGAAATGGAATGAGGTTCACTTC 840  
Db |||||  
2536 GCTTCAGTAACCTCTGAGCTTGGATACCTTTGGCGGAGAAATGGAATGAGGTTCACTTC 2595

RESULT 2  
AF082744 11329 bp DNA linear BCT 18-SEP-2000  
LOCUS  
DEFINITION Ehrlichia canis p28 multigene locus, partial sequence.  
ACCESSION AF082744 AF168788 AF168789  
VERSION AF082744.2 GI:10181081  
KEYWORDS  
SOURCE Ehrlichia canis.  
ORGANISM Ehrlichia canis  
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
Anaplasmataceae; Ehrlichia.  
REFERENCE  
1 (bases 1 to 11329)  
McBride, J.W., Yu, X.J. and Walker, D.H.  
Molecular cloning of the gene for a conserved major immunoreactive  
28-kilodalton protein of Ehrlichia canis: a potential  
serodiagnostic antigen  
Clin. Diagn. Lab. Immunol. 6 (3), 392-399 (1999)  
JOURNAL  
MEDLINE 99242757  
PUBMED 10225842  
REFERENCE  
2 (bases 1 to 11329)  
McBride, J.W., Yu, X.J. and Walker, D.H.  
A conserved, transcriptionally active p28 multigene locus of  
Ehrlichia canis  
Gene 254 (1-2), 245-252 (2000)  
JOURNAL  
MEDLINE 20432107  
PUBMED 10974556  
REFERENCE  
3 (bases 1 to 11329)  
McBride, J.W., Yu, X.J. and Walker, D.H.  
Direct Submission  
Submitted (07-AUG-1998) Pathology, University of Texas Medical  
Branch, 301 University Blvd., Galveston, TX 77555-0609, USA  
4 (bases 1 to 11329)  
McBride, J.W., Yu, X.J. and Walker, D.H.  
Direct Submission  
Submitted (04-AUG-2000) Pathology, University of Texas Medical  
Branch, 301 University Blvd., Galveston, TX 77555-0609, USA  
REMARK  
Sequence update by submitter  
COMMENT On Sep 18, 2000 this sequence version replaced gi:3769522.  
FEATURES  
source  
1..11329  
/organism="Ehrlichia canis"  
/strain="Juke"  
/db\_xref="taxon:944"  
<1..146  
/gene="p28a"  
<1..146  
/gene="p28a"

/note="outer membrane protein"  
/codon\_start=3  
/transl\_table=11  
/product="p28"  
/protein\_id="AAG14355.1"  
/db\_xref="GI:10181083"  
/translation="SIISHSVKRLPVQYNTLEEYPRVTSATLIDIGYLGSGIGIR  
FIF"  
170..1051  
/gene="p28-1"  
170..1051  
/gene="p28-1"  
/note="outer membrane protein"  
/codon\_start=1  
/transl\_table=11  
/product="p28-1"  
/protein\_id="AAG14356.1"  
/db\_xref="GI:10181084"  
/translation="MNNKLTITNTVLCLSLPNISSKAINNNKYYGLYISGQ  
YKPSVSVSNFVKETNVTIKNLIALKKVDISIETKDSVSGISNPSNFTIPYAVFQ  
DKSVNFGTITGYTFAEGTRVEIEGSEVEFDVKNPGGYTLDAYRYFALAREMKGNSFT  
PKKVNSTIPIHTVMRDGLSIISVIVNVCVDFSLNNLSISPYICGGAGVDAIBPFDVL  
HIKFAVQSLGIAYSLPSNISLFSALYYHKVMGNQPKNLNVQHVAEIASIPKITSAVA  
TLNIGFPGSIGARLTF"  
1599..2441  
/gene="p28-2"  
1599..2441  
/gene="p28-2"  
/note="outer membrane protein"  
/codon\_start=1  
/transl\_table=11  
/product="p28-2"  
/protein\_id="AAG14357.1"  
/db\_xref="GI:10181085"  
/translation="MNVKLTIVLSLISLIPYOSFADPVGSRTNDNKEGFYISAK  
YFSGISHPKFSAEETPINGTNSLTKKVGLKGDITKDDFTRVAPGIDFQNNLIS  
GFSGISGMDGPRIELEAAAYQQFNPKNNTDNDNGEYKHFALSRRDAMEDQQYVV  
LKNDDITFMSLVNTCYDITAEGVFVPYACAGIGADLITFKDLNLKFPAYQKGIGIS  
YPTTPEVSAGFVGYYHGVIGNKEKIPVITPVVLNDAPQTTSASVTILDVGVFGSEIGM  
RFTF"  
2769..3599  
/gene="p28-3"  
2769..3599  
/gene="p28-3"  
/note="outer membrane protein"  
/codon\_start=1  
/transl\_table=11  
/product="p28-3"  
/protein\_id="AAG14358.1"  
/db\_xref="GI:10181086"  
/translation="MNVKLTIVLSLISLIPYOSFADPVGSRTNDNKEGFYISAK  
PSISHFGNFSKAEKNTTGI FGLKESWTGGIILDKHEAAPTNPYGFYKNNPFLGF  
AGVIGYISGSPRIEFSEVETFDVQNGKFNNDHAKYCALSDSSKTMESGKFLPK  
NEGLDISLNLVYDILNKMPFSPYICAGIGTDLIFMFDALNHQAAVYQKLGFTNP  
ISPEANISMGVHFKVTNNEFRVPLLTAGGLAPDNLFAIVKLSICHFGLFEGYRVSP  
"  
3810..4673  
/gene="p28-4"  
3810..4673  
/gene="p28-4"  
/note="outer membrane protein"  
/codon\_start=1  
/transl\_table=11  
/product="p28-4"  
/protein\_id="AAG14359.1"  
/db\_xref="GI:10181087"  
/translation="MKYKTFVTVALVLTSTFTHFPFYPSPARASTIHNFYIGKYM  
TASHFGIFSAKEQSFTKVLGLDQRLSHNIINNNDTAKSLKQVNSFKYKNNPFLGF  
AGAIGYISGNSRIEVSVEI FDTKPNGNLNDSHKYCALSHGSHICSDGNSGDWYT  
AKTDKFLVLLKNEGLDVSFNLNACVDTTEKMPSPYICAGIGTDLISMPETQNKIS  
YQKGLGNTYNSRVSVFAGGHFKVIGNEFKIGIPILLPDGSNIKVQQSATVILDVCH  
FGLIEGSRFFP"  
4970..5821

gene

CDS





```

/transl_table=11
/product="major outer membrane protein P30-14"
/protein_id="AAK28686.1"
/db_xref="GI:13512592"
/translation="MQRLNFINIIAFLALLLPQSFSLYIHDHEITQNGVGLYISSQ
YKSIYFKNFLIEBNSHKTEVLMLGLANDVHTEYVLKONTKFNTPYSAKPNLSLIN
LSGAIYSGQGPLRIEISYENGVASCKCNPKVNIYIALVRDKPGNITPDQHS
LNSMYTITKNGIGISLVINGCYDIAFSNVKISPVYICAGIGGDFITFLFTMTHKF
AYQKFGISLYLVSPTSISIFANGHYHKVMONVFNKLHVKIYIKLQDAPTITTSARAKLRI
GYGFSVGVRFVP"
gene
7419. 8255
/gene="p30-13"
CDS
7419. 8255
/gene="p30-13"
/notes="p30 family member"
/codon_start=1
/transl_table=11
/product="major outer membrane protein P30-13"
/protein_id="AAK28687.1"
/db_xref="GI:13512593"
/translation="MNNKSLIIGTILLFSSLPIKAFSVINHSDISSNVNGLVYPTG
QYRPVSHSGFVRETNIATQOLVSLNTKNENHIIERTNFSGIYAKFQNAASFS
GATGYSYPEGLKEPEIEISYEKFGVKSTQSNVAVI FALARTTSSNPSNDKNVYVMK
NSGLSVASWINGCYNMSFNLVSPYICAGIDGIEFPDFTLYIKLAVQSKLGVNYS
LSRFRNIFADMYTHKVIQNGQFNLVHIAVALDFFPKVTSATILNLVAIFGGEVGIRI
IL"
CDS
8266. 8823
/codon_start=1
/transl_table=11
/product="unknown function U3"
/protein_id="AAK28688.1"
/db_xref="GI:13512594"
/translation="MATYMIHKAKIIEPMFSPSILTDYIILSLNNVDHNSFFIICT
KLYKLILSSITKKLINVKONIAVLHINILYLNKCSKFKSTQSFVEMVFTYTKNIKDR
ILNFNSLYFTKYNVTKNSNESYKSTYKSHNFENATYTTSTFIHQHKLVTYSITNQH
NKSTKSPNKKILAKNILNQNSEYE"
gene
8816. 9685
/gene="p30-12"
CDS
8816. 9685
/gene="p30-12"
/notes="p30 family member"
/codon_start=1
/transl_table=11
/product="major outer membrane protein P30-12"
/protein_id="AAK28689.1"
/db_xref="GI:13512595"
/translation="MNSKTFSLIGLSILICLAACLPQGFSSSNVYNTKHTGLYIS
GLYKPSVSHSDFSIKETIYNTNTEALFGLKODISSILRNKFTQYNNFNVPYTAKPFD
DFASFTAVGYIANGPRIIEGYSVEEDVKNPGNYTTIDAHYIALAREKYSYVLS
PKENKXVYIKNGISIVSIILINGCYDISLNSDKVSPYICGFGDGFIEFSAIRFKFA
YQKIGISLYSLNIIILFTDGYVHKVINSQFNLVNHEVNVNELTTPDKVTSATAFILNI
EYFGGFGFLKFI"
Query Match 100.0%; Score 840; DB 1; Length 28254;
Best Local Similarity 100.0%; Pred. No. 1.4e-140;
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATCAATTATAAGAAAAATCTTAGTAAGAAGCGCGTTAATCTCATTAATGTCAATCTTACCA 60
Db 15562 ATCAATTATAAGAAAAATCTTAGTAAGAAGCGCGTTAATCTCATTAATGTCAATCTTACCA 15622
Qy 61 TATCAGTCTTTTCAGATCTCTGAGGTTCAAGACTAATCATACAAAGAGGCTTCTAC 120
Db 15622 TATCAGTCTTTTCAGATCTCTGAGGTTCAAGAACTAATGATACAAAGAGGCTTCTAC 15682
Qy 121 ATTAGTGCAAGTACAAATCCAAGTATATCACCTTTAGAAAAATCTCTGCTGAAGAAACT 180
Db 15682 ATTAGTGCAAGTACAAATCCAAGTATATCACCTTTAGAAAAATCTCTGCTGAAGAAACT 15744
Qy 181 CCTATTAAATGGACAAATTCCTCCTACATAAAAAAGTTTTTCGGACATAAAGAAAGATGGTAT 240
Db 15742 CCTATTAAATGGACAAATTCCTCCTACATAAAAAAGTTTTTCGGACATAAAGAAAGATGGTAT 15802

```

USA  
4 (bases 1 to 27190)  
REFERENCE  
AUTHORS Chashi,N., Zhi,N., Zhang,Y. and Rikihisa,Y.  
TITLE Direct Submission  
JOURNAL Submitted (27-AUG-1997) Department of Veterinary Biosciences, The Ohio State University, 1925 Coffey Rd., Columbus, OH 43210-1093,  
USA  
5 (bases 1 to 27190)  
REFERENCE  
AUTHORS Chashi,N., Rikihisa,Y. and Unver,A.  
TITLE Direct Submission  
JOURNAL Submitted (29-NOV-2000) Department of Veterinary Biosciences, The Ohio State University, 1925 Coffey Rd., Columbus, OH 43210-1093,  
USA  
REMARK  
COMMENT Sequence update by submitter  
On or before Apr 2, 2001 this sequence version replaced gi:2853584, gi:2853273.  
FEATURES  
source  
1. 27190  
Location/Qualifiers  
/organism="Ehrlichia chaffeensis"  
/strain="Arkansas"  
/db\_xref="taxon:945"  
315..956  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical transcriptional regulator"  
/protein\_id="AAC02808.2"  
/db\_xref="GI:13511828"  
/translation="MSTHAKNINKMEDTDEIKAHDPDGLIAGLIDKNQKKNFTNKGYN  
DDHTSKNPTKEGVSTDKLKNSTTKIRPHVPDECVGKEIKRQIRMRGMSQNLAKLIG  
ITFOOVQYKEGTNRIVLSRLYQLASVLNVEVRDMLKLQEDLKNISCDHTVASTSL  
KDNEEKIPEFSDHSKIDSKVELMVMRAYTCTIKNEKVRNIITNLKVALSLONQK"  
1136..2032  
/gene="omp-1M"  
1136..2032  
/gene="omp-1M"  
/note="omp-1M"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1M"  
/protein\_id="AAK28658.1"  
/db\_xref="GI:13511829"  
/translation="MILINMKLFYHLONIMTKPSAIGIVLSLTLFACNVFASPID  
FNSSETAGFVASHYNIQPPRPSIVKKTIDENTERKELFTLKESTDTPTKQKS  
EFNDKKGSPILNRYNIGFSGAIGSGGLRLEAFPDVDKQYKKNYFAL  
CKDQSIESTDMSNGHVMKNEGRFVSLTFNACYDMIVSNSSILVPSACIGIGQITN  
PLGNTNITLTKAKLGLGLFSLSPKTVIFANGYYVTKTSDNSPTLSVQVPLELKEAPKH  
IDPIACFNADNYGGEVGLRFTL"  
2498..3352  
/gene="omp-1N"  
2498..3352  
/gene="omp-1N"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1N"  
/protein\_id="AAK28659.1"  
/db\_xref="GI:13511830"  
/translation="MSKRSNRKFLVWMLILFTPHISLASVLNDHNSMYVGIQYKPAR  
OHLKSLIKESAANTVEVGLKDLLNDLLTGIKDNTFNKIKYPPYNNRNLGSGIP  
GYNNKFRIESELSYETPHIKNGYKRIDCEKHPALAKEISGGSNNPANNKYVTLIN  
NGISLTALINVCYDLDGRHNIITYISCLGFGVDTIDLSKYITTKFSGYQKLGASYV  
SPQSVFTIEGYHGLFGKKEKIPVNPYPCDPPSPTPNSKPHVHTTALAMLSIGYYGG  
SIGIKFIL"  
3459..3905  
/codon\_start=1  
/transl\_table=11  
/product="unknown function UN2"  
/protein\_id="AAK28660.1"  
/db\_xref="GI:13511831"  
/translation="MNNKSRGQVLLFAAYSIPKYNISCTSKFRLSVKCFVYNPLKYF  
YKSTFLKGLGKFFVNNYSKTSNVLHSLRFLKGLIITNPKISLYLNFFTNPRVLIY  
YKLNKCNSSINLYLNKNTIINTATISFINSVLIIGYPGVNTNY"  
3915..4808  
gene

CDS  
/gene="omp-1Q"  
3915..4808  
/gene="omp-1Q"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1Q"  
/protein\_id="AAK28661.1"  
/db\_xref="GI:13511832"  
/translation="MSYAKVFILICLLIVPSLSFAIVNNDFLKDNIGHFVIGQYKPK  
GVPRNFPLVTNNIRLMSDEECSTIPHMVQSVAGTLPPEALRELAKGLHGGY  
LATFLPNPYKKNLLCAGGVIGYSTTHFRVEVEAYEKENLTPAGYLNKPYEYFA  
LATMDTKHQHQAEDKIYMKNTGITLSFIIINACYDFILKTRNVAPLYCLGVGGN  
FIPLDQVSKPAQYQAKVIGISFVSPNIAFFIDGSEFHGLNNOFSLPVDVSSSGFP  
TISAKFNANFTSSIGIRFIS"  
4825..5682  
/gene="omp-1P"  
4825..5682  
/gene="omp-1P"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1P"  
/protein\_id="AAK28662.1"  
/db\_xref="GI:13511833"  
/translation="MQKLYISFILISGLLPLKYVFCMHQNNNIDGSYVTIKYQLTTPH  
FQPFYIKETDFTQEPICGLAKITANTKFTDLKENFSPLHQTDSYKSYQNDLLGIGL  
SVGLFVKSPRIEFEGAVKSKDGKYKYPFAPRKSBSHGLDNTFGYTVAK  
NNGISITSNILINLCSETKYKAPTYICIGVGGFIELEFDMRIKEAQQKVGVSYPIT  
SKILSINGQYHKVIGNKFFELLPYQPVVELKRLVTNTKSDIDQDVTASLTNLNDEHS  
SEIGLSFIF"  
5695..6513  
/gene="omp-1T"  
5695..6513  
/gene="omp-1T"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1T"  
/protein\_id="AAK28663.1"  
/db\_xref="GI:13511834"  
/translation="MYNNKHYCIYVYVITLFFLLPLIESLALIGNVEKDLKVSS  
TVSSQYKPSIFHRNPSIOEHPKSSSEEFKKIKANLNILKSNALNQLQDNTSF  
SGTGYSGKRLREAEQCYEFNVKSNNSLIISNKYHSRIHDENAITNNKLSIA  
SIMWTYCI SINNTSIVPLCTGIGEDLVGLFNTIHFKLAYQCKGVMSYLIINNILL  
FDIYVHKVMGNRFKNLYMYVADPNISEITPILAKLDIGYFGSEIGIRMFEN"  
6529..7416  
/gene="omp-1U"  
6529..7416  
/gene="omp-1U"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1U"  
/protein\_id="AAK28664.1"  
/db\_xref="GI:13511835"  
/translation="MTKKFNFNVLFTLLFLPLKSFYVANNNTITQKGLYISGO  
YKPSIFHKVFSVEENDKVDLIGLTDVTYITEHLRDNTKPNTHYIAKPKNNFINF  
SSAIGYSGOGPLEIESSYGDENVYKNYAVQDVNRYFALVRKNGSNPSPKPHET  
SQSDSNPKSPFTLMKNNGVFNVASVIINGCYDFSNNTTISPVCICVGGDPIEFPE  
VMIHKACQSKGVISYPISESIITIFADAHYHKVINNNFNHLVKSVELKNSPTTISA  
TAKLNIEYFGGEGMRPIF"  
7438..8277  
/gene="omp-1V"  
7438..8277  
/gene="omp-1V"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1V"  
/protein\_id="AAK28665.1"  
/db\_xref="GI:13511836"  
/translation="MSKKFITTITGLASLLSLFSLIESFSGAINHNHNTGNTSGIYITG

```
QVRPGVSHFSNFVKTNDVTIQLVGYKKSSSIDPNTYSNFQGYPTVTTFODNAASF5
GAIQSVESIRLEEGSEYEDVDKPKDYSAKDAFRFALARNITSTVPDAQXYTM
KNGLSVASINGWCYDLSFNNLVSPYICAGIGEDFIEFDTHLIKLAQCKLGISY
YFFPKINVFAGYHVRVIGNFKFNLNVNHHVTLDEFFPKATSAVATLNVAYFGGAGVK
FTF"
CDS
8300..8890
/codon_start=1
/transl_table=11
/product="unknown function UN3"
/protein_id="AAK28666.1"
/db_xref="GI:13511837"
/translation="MIYKEKLRVGEYILAYLSFYLVLNIRYNSLAICVI
SLRTNINVSSTKKLIKDKCRDFKSNMNCVLYGKPLNLOIYFGIFSRFONNTLI
IPNDSKGFYTLWNPALHYTYLTJGSEYENFDILYENIICQCKLLINRSLVING
HNKNTLVIIIPNAREFSNEIRVNSINKESSYEC"
8880..9731
/gene="omp-1w"
/gene="omp-1w"
/feature="omp-1 family member"
/codon_start=1
/transl_table=11
/product="major outer membrane protein OMP-1w"
/protein_id="AAK28667.1"
Query Match 71.5%; Score 600.2; DB 1; Length 27190;
Best Local Similarity 82.7%; Pred. No. 78-98;
Matches 702; Conservative 0; Mismatches 138; Indels 9; Gaps 1;
QY 1 ATGAATTATAAGAAAATCTAGTAAAGACGCGTAAATCTCAATATGTCATCTTACCA 60
DB 15670 ATGAATTACAGAAAATTTTGTAGCAGTGATTAATTTCAATATGTCATCTTACT 15729
QY 61 TATCAGTCTTTGCGAGATCCTGTAGGTCAA-----GAACTAATGATTAACAGAA 111
DB 15730 TACCAATCTTTGCGAGATCCTGTAACTTCAATGATACAGGAATCAACGACAGAGAA 15789
QY 112 GCCTTCTACATGAGCAAGTACATCCAGTATATCACATTTAGAAAATTCCTCTCT 171
DB 15790 GCCTTCTACATGAGTAAAGTATATCCAGCATATACATCTCAGAAAATTCCTCAGT 15849
QY 172 GAAGAACTCTTATTAATGGAACAAATCTCTCACTAAAAAGTTTTCGGACTAAAGAA 231
DB 15850 GAAGAACTCTTATTAATGGAACAAATCTCTCACTAAAAAGTTTTCGGCTGAAAAA 15909
QY 232 GATGTGATATTAACAAAAAGACGATTTTCAAGAGTAGTCCAGGCAATGATTTTCAA 291
DB 15910 GACGAGATATAGCAAAATCTCGAAATTTTAAACAGGACAGATCCAGCCCTCGAGTTTCAG 15969
QY 292 AATAACTTAAATACAGGATTTTTCAGGAGTATTTGTTACTCTATCGAGCGACCAAGATA 351
DB 15970 AATAACCTTAAATACAGGATTTTTCAGGAGTATTTGTTACTCTATCGAGCGACCAAGATA 16029
QY 352 GAACTTGAAGCTGCATATCAACAATTTTAAATCCAAAAAACACCGATAACAAATGATCTGAT 411
DB 16030 GAACTTGAAGCTGCATATCAACAATTTTGAATGCAAAAAATCTTGACACAAATGACACTAAT 16089
QY 412 AATGTGAAATATATAAACAATTTTGCATTAATCTGTAAGATGCAATGGAAGATCAGCAA 471
DB 16090 AGCGGTGACTACTATAAAATCTTTGGACTATCTCGTGAAGAGCAATAGCAGATAAGAAA 16149
QY 472 TATGTAGTACTTAAATAGCGGCACTATTTTATGTCATTTGCTATGTTAATCTTCTAT 531
DB 16150 TATGTGTGCTTAAATAGCGGCACTATTTTATGTCATTTGCTATGTTAATCTTCTAT 16209
QY 532 GACATTACAGCTGAAGGAGTATCTTTTCGTACCATATGATGTCAGGATAGGAGCAGAT 591
DB 16210 GACATTACAGCTGAAGGAGTATCTTTTCATACCGTATGATGTCAGGATAGGAGCAGAC 16269
QY 592 CTTATCACTATTTTAAAGACCTCAATCTAAATTTTGGCTTACACAGGAAAAATAGGTATT 651
DB 16270 CTTATAAAGCTATTTAAGGATTTTAAATTTTAAATTTTCTATACCAAGGAAAAATAGGTATT 16329
```

```
QY 652 AGTTACCTATCACACAGAGTCTCTGCATTTTATGTTGATACTACCATGGCGTTATT 711
DB 16330 AGCTATCCATCACACAGAGTTCGCTTTTATGAGGATACTACCATGGAGTTATA 16389
QY 712 GGTAAATAATTGAGAGATACCTGTAAATACTCTCTAGTATTAATGATGCTCTCTCAA 771
DB 16390 GGAATAATAATTGAGAGATACCTGTAAATACTCTCTAGTATTAATGAGAGTCTCTCTCAA 16449
QY 772 ACCACATCTGCTTCAGTAACTCTTGAGCTTGATCTTTGGCGAGAAATTTGGAATGAGG 831
DB 16450 ACCACATCTGCTTCAGTAACTCTTGAGCTTGATCTTTGGCGAGAAATTTGGAATGAGG 16509
QY 832 TTCACCTTC 840
DB 16510 TTCACCTTC 16518
RESULT 5
AF125276 3507 bp DNA linear BCT 20-APR-1999
LOCUS Cowdria ruminantium isolate Highway major antigenic protein 1 like
DEFINITION protein and major antigenic protein 1 (MAP1) genes, complete cds.
ACCESSION AF125276
VERSION AF125276.1 GI:4589100
KEYWORDS Ehrlichia ruminantium.
SOURCE Ehrlichia ruminantium.
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
REFERENCE 1 (bases 1 to 3507)
AUTHORS Sulsona,C.R., Mahan,S.M. and Barbet,A.F.
TITLE The map1 Gene of Cowdria ruminantium Is a Member of a Multigene
JOURNAL Family Containing Both Conserved and Variable Genes
REFERENCE Biochem. Biophys. Res. Commun. 257 (2), 300-305 (1999)
AUTHORS Sulsona,C.R., Mahan,S.M. and Barbet,A.F.
JOURNAL Direct Submission
TITLE Submitted (02-FEB-1999) College of Veterinary Medicine Department
of Pathobiology, University of Florida, P.O. Box 110880,
Gainesville, FL 32610, USA
FEATURES
Location/Qualifiers
1. 3507
/organism="Ehrlichia ruminantium"
/isolate="Highway"
/db_xref="taxon:779"
948..1796
/note="ORF2"
/codon_start=1
/transl_table=11
/product="major antigenic protein 1 like protein"
/protein_id="AAD26347.1"
/db_xref="GI:4589101"
/translation="MNYKKILVRSALISLMSFLPYQFAEPVPSNNINENAKSGFYI
SAKYNPIPHRFKESABETPVYKDSITPKVGLKSGSITKYSDFTRTDSIFGQNN
FISGSGSIGYIMDPRVEIAAYQKFNPKANETDTSYKHYKHSRASTMTDKKY
VLTNNGVTFSSLMFNACYDITAEQVDFIPVACAGIGADLLISIFDDINLKAYQCKIG
ISYPTFEISAFIGYHGVGNKXKVPKLPVTLTADAPQSTASVTLTADAGYFGGEL
GVRTTF"
2155..3018
/gene="MAP1"
2155..3018
/gene="MAP1"
/codon_start=1
/transl_table=11
/product="major antigenic protein 1"
/protein_id="AAD26348.1"
/db_xref="GI:4589102"
/translation="MNCKIPIFTSLISLVSLFQVPSFSDVIQDSNPAQSVYISAKY
MPTASHFGKASIKEDSKNTQVGLKKMDGKVTSPSSNTNSTIFTKDYSPRYENN
PFLGAGAGISGMMGPRIEFVETFDVRKPNPGKNYKNNHMYCALDTAAQNSNGAG
LTTVMVKNENLTNLSLMNACYDILMDGIPVSPVCAGIGTDLTDLVSLNATNPKLSYQ
KLGISYSINSEASIFIGGHPHRVIGNEFKDIALTKIFTSKTGISNPGFASATLDVCH
FGIEIGGRFVF"
```



BASE COUNT	1288 a	473 c	459 g	1287 t					
ORIGIN									
Query Match	62.9%;	Score 528.2;	DB 1;	Length 3507;					
Best Local Similarity	78.2%;	Pred. No. 8.8e-85;							
Matches	664;	Conservative	0;	Mismatches 173;	Indels 12; Gaps 2;				
Qy	1	ATGAATTAAGAAAAATCTAGTAAGAGCGGTTAATCTCAATTAATGCTCAATCTTACCA	60						
Db	948	ATGAATTAAGAAAAATCTCGTAGAAGTGGTTAATCTCAATTAATGCTCAATCTTACCA	1007						
Qy	61	TATCAGTCTTTTGCAGATCTGTAGGTTCAA-----GAACCTAATGATAACAAGAA	111						
Db	1008	TATCAGTCTTTTGCAGAACCTGTAGTTCAATTAACATTTGGAATGAAATGCTTAAGAA	1067						
Qy	112	GGCTTCTACATTAGTGCAAGTACAAATCAAGTATATCACACTTTTAGAAAAATCTCTGCT	171						
Db	1068	GGGTTCTACATAAGTGCAAAATACAAACCAAGCATACCACACTTCAGAAAAATTTCTGCT	1127						
Qy	172	GAGAACTCTTATTAATGGAACAAATCTCTCACTAAAAAGTTTTCGGACTAAAGAA	231						
Db	1128	GAGAACTCTGTATACGGTAAAGACTCTCCAACTAAAAAGGTATTGTGGTTAAAAAAG	1187						
Qy	232	GATGGTATATACAAAAAGACGATTTTACAGAGTAGTCCAGGCAATTTGATTTCAA	291						
Db	1188	GAGGGTTCTATAACAAAAATACAGTATTTCTACTAGAACAGATATATCGTTTGGGGCCAA	1247						
Qy	292	AATAACTTAATATCAGGATTTTTCAGGAAGTATTGGTTACTCTATGACGACCAAGATA	351						
Db	1248	AATAATTTTATCTCAGGTTTCTCAGGAAGCATAGTTTATATCATGGATGGACCAAGTA	1307						
Qy	352	GAACCTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGATAACAATGATCTGAT	411						
Db	1308	GAGATTGAAGCTGCATACCAAAAAATTCACCCAAAAAATCCAGCT---AATGAACTGAT	1364						
Qy	412	AATGGTGAATACTATAACAATTTTTCGATTTATCTGTAAGATGCAATGGAATCAGCAA	471						
Db	1365	ACTAGTGATTACTATAAACACTATGGATTATCTCGTGCAAGAACCATGACAGATAAAAA	1424						
Qy	472	TATGTAGTACTTAAAAATGACGCATACCTTTTATGTCTATGATGGTTAATCTTGCTAT	531						
Db	1425	TATGTGTACTTACAATAATGAGTAACCTTTTTCATCATTAATGTTTAACTGCTGCTAT	1484						
Qy	532	GACATTACGCTGAAGGATATCTTTCTGATACCATATGATGATGAGGATPAGGACGAT	591						
Db	1485	GACATTACAGAGAAGGATACCTTTTCATTCATATGATGATGCTGCTGTTTGGTGCGAT	1544						
Qy	592	CTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGAAAAATAGGTATT	651						
Db	1545	CTTATATCTATATTGTATGATATAAATTTAAAAATTTGCTTACCAAGGTAATAATTTGGTATT	1604						
Qy	652	AGTTACCTTATCACACAGAGTCTCTGCAATTTTATTTGGTGATACCTACCATGGCGTTATT	711						
Db	1605	AGTTATCCAATTAATCTCTGAAATTTCTGCAATTTTATTTGGCGGATATTACCATGGAGTAATA	1664						
Qy	712	GGTAATAAATTTGAGAAGATACCTGTATAACTCTGTAGTATTAAATGATGCTCTCTCAA	771						
Db	1665	GGTAACAAGTATAACAAAGTACCTGTAAAGCTTCTCTGTAACCTTTTACAGATGCTCTCTCAA	1724						
Qy	772	ACCACTCTGCTTCAAGTAACTCTTGACGTTTGGATATCTTTTGGCGGAGAAATTTGGAATGAGG	831						
Db	1725	AGCACTTCCGCTTCAGTAACTCTTGAACGCTGGATATTTTGGTGGTGAATCTTGGAGTAAG	1784						
Qy	832	TTACACCTTC 840							
Db	1785	TTTACTTTC 1793							
RESULT 6									
AF125274	3535 bp	DNA	linear	BCT 20-APR-1999					
LOCUS									
DEFINITION	Cowdria ruminantium isolate Weigevonden major antigenic protein 1 like protein and major antigenic protein 1 (MAP1) genes, complete								
cds.									
AF125274									
AF125274.1	GI:4589094								
KEYWORDS									
SOURCE									
ORGANISM	Ehrlichia ruminantium.								
	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;								
	Ehrlichia ruminantium.								
	Anaplasmataceae; Ehrlichia.								
REFERENCE									
AUTHORS	Sulsona, C.R., Mahan, S.M. and Barbet, A.F.								
TITLE	The map1 Gene of Cowdria ruminantium Is a Member of a Multigene Family Containing Both Conserved and Variable Genes								
JOURNAL	Biochem. Biophys. Res. Commun. 257 (2), 300-305 (1999)								
REFERENCE									
AUTHORS	Sulsona, C.R., Mahan, S.M. and Barbet, A.F.								
TITLE	Direct Submission								
JOURNAL	Submitted (02-FEB-1999) College of Veterinary Medicine Department of Pathobiology, University of Florida, P.O. Box 110880, Gainesville, FL 32610, USA								
FEATURES									
source	1. 3535								
	/organism="Ehrlichia ruminantium"								
	/isolate="Weigevonden"								
	/db_xref="taxon:779"								
	987..1835								
	/note="ORF2"								
	/codon_start=1								
	/transl_table=11								
	/product="major antigenic protein 1 like protein"								
	/protein_id="AAD26343.1"								
	/db_xref="GI:4589095"								
	/translation="MNYKKILVRSALISLMSFLPYQSPABVPVSNINENAKGFIY								
	SAKYPGSIPIHPRKFSABETPVKDSPTKKVFGKKEGSIKYSDFTDTSFQGN								
	FISGSGSIGVIMDPVRVEIAAYOKFNKPANETDTSYKHVGLSRAMADKKY								
	VLTNGVTFSLSMKNACYDTAGVFPFYPACAGICADLISIEDINLKFAVQKGIG								
	ISYPTPEISAFIGGYTHGVGNKINPVKLPVLTIDAPQSTASVTLDAIGYFGGEL								
	GVRFPT"								
gene	2210..3055								
	/gene="MAP1"								
	2210..3055								
	/gene="MAP1"								
	/codon_start=1								
	/transl_table=11								
	/product="major antigenic protein 1"								
	/protein_id="AAD26344.1"								
	/db_xref="GI:4589096"								
	/translation="MNCRKIFITSLISLVFLPGVSFSDAIQEDSSPVGSVVISAKY								
	MPTASHFGKMSIKEDSRDTKVFGKKDGVKTSSTNITFEKDYSPKYNPNPFLGP								
	AGAIGYSMNPGRIEFSEISYETFDVKNPGNYKNDAMHYCALDTSFGGAAASVMV								
	KNENLTDISLMLNACYDMLDGMPPVPCAGIGTDLVSNITNPKLSYQKGLGISY								
	SINPEASIFIGGHFHRVIGNEFKDITTSKIFNTSGATPGFASATLDVCHFGIEIG								
	GRFVF"								
BASE COUNT	1290 a	445 c	480 g	1320 t					
ORIGIN									
Query Match	62.9%;	Score 528.2;	DB 1;	Length 3535;					
Best Local Similarity	78.2%;	Pred. No. 8.7e-85;							
Matches	664;	Conservative	0;	Mismatches 173;	Indels 12; Gaps 2;				
Qy	1	ATGAATTAAGAAAAATCTAGTAAGAGCGGTTAATCTCAATTAATGCTCAATCTTACCA	60						
Db	987	ATGAATTAAGAAAAATCTCGTAAGAGTGGGTTAATCTCGTTAATGCTCAATCTTACCA	1046						
Qy	61	TATCAGTCTTTTGCAGATCTGTAGGTTCAA-----GAACCTAATGATAACAAGAA	111						
Db	1047	TATCAGTCTTTTGCAGAACCTGTAGTTCAATTAACATTTGGAATGAAATGCTTAAGAA	1106						
Qy	112	GGCTTCTACATTAGTGCAAGTACAAATCAAGTATATCACACTTTTAGAAAAATTTCTGCT	171						
Db	1107	GGGTTCTACATAAGTGCAAAATACAAACCAAGCATACCACACTTCAGAAAAATTTCTGCT	1166						
Qy	172	GAGAACTCTTATTAATGGAACAAATTTCTCACTAAAAAGTTTTCGGACTAAAGAA	231						

```
Db 1167 GAGGAACCTCCTGTATACGGTAAAGACTCTCCAACTAAAAAGGTATTTGGGTAAAAAAG 1226
Qy 232 CATGGTGTATTAACAAAAAGACAGATTTTACAAGAGTAGTCCAGGCAATTGATTTTCAA 291
Db 1227 GAGGTTCTTAACAAAAATACAGTGAATTTCTACAGACAGATATATCGTTTGAGGGCCAA 1286
Qy 292 AATAACTTAATACAGGATTTTCAGGAAGTATTGGTTACTCTATCGAGCGGACCAAGATA 351
Db 1287 AATAATTTTATCTCAGGTTCTCAGGAAGCATAGGTTATATCATGATGGACCAAGATA 1346
Qy 352 GAACCTGAAGCTGCATATCAACAATTTTAAATCCAAAAAACACCGATAACAATGATCTGAT 411
Db 1347 GAGATTGAAGCTGCATACCAAAAATTCACCCAAAAATCCAGCT--AATGAAACTGAT 1403
Qy 412 AATGGTGAATACTATAAACAATTTTCATATCTCTGTAAGATGCAATGAAGATCAGCAA 471
Db 1404 ACTAGTGATTACTATAAACAATGATGATTTCTCGTGAGAAGCTATGCGCAGATAAAAAA 1463
Qy 472 TATGTAGTACTTAAAAATGAGGCATAACTTTTATGTCATTGATGTTAATACTTTGCTAT 531
Db 1464 TATGTGTACTTACAAATAATGGAGTAACCTTTTTCATTAATGTTCAATGCTTGCTAT 1523
Qy 532 GACATTACAGCTGAAGAGTATCTTTTCGTACCATATGATGTCAGGATATAGGACAGAT 591
Db 1524 GACATTACAGCAGAGGAGTACCTTTTCATTCATATGATGCTGCTGTTATTTGGTCAGAT 1583
Qy 592 CTTATCATTATTTTAAAGACCTCAATCTAAAATTTGCTTACCAAGGAAAAATAGGTATT 651
Db 1584 CTTATATCTATATTTGATGATATAAATTTAAAATTTGCTTACCAAGGTAATAATTTGAT 1643
Qy 652 AGTTACCTCTACACACAGAGTCTCTGATTTATTTGTTGGTACTACCATGGCTGTTATT 711
Db 1644 AGTTATCCAATTAATCTCTGAAATTTCTGCAATTTATTTGGGATATTAACATGGAGTAATA 1703
Qy 712 GGTAAATAATTTGAGAAGATACCTGTAATAACTCTCTGATGATTAATTAATGATGCTCTCAA 771
Db 1704 GGTAAACAGTATAACAAAGTACCTGTAAGCTTCTGTAACCTTTTAAACAGATGCTCTCAA 1763
Qy 772 ACCACATCTGTTAGTAACTCTTGACGTTGATGATCTTTGGCGGAGAAATTTGGAATGAGG 831
Db 1764 AGCACTTCGCTTCAGTAACTCTTGACGCTGGATATTTTGGTGGTGAACCTTTGGAGTAAGG 1823
Qy 832 TTCACCTTC 840
Db 1824 TTTACTTTC 1832
```

```
RESULT 7
AF125277
LOCUS AF125277 3538 bp DNA linear BCT 20-APR-1999
DEFINITION Cowdria ruminantium isolate 'Lemcot3' major antigenic protein 1 like
protein and major antigenic protein 1 (MAP1) genes, complete cds.
ACCESSION AF125277
VERSION AF125277.1 GI:4589103
KEYWORDS
SOURCE
ORGANISM Ehrlichia ruminantium.
Ehrlichia ruminantium
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE
AUTHORS Sulsona,C.R., Mahan,S.M. and Barbet,A.F.
TITLE The map1 Gene of Cowdria ruminantium Is a Member of a Multigene
Family Containing Both Conserved and Variable Genes
JOURNAL Biochem. Biophys. Res. Commun. 257 (2), 300-305 (1999)
REFERENCE
AUTHORS Sulsona,C.R., Mahan,S.M. and Barbet,A.F.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-1999) College of Veterinary Medicine Department
of Pathobiology, University of Florida, P.O. Box 110880,
Gainesville, FL 32610, USA
FEATURES
source
1..3538
/organism="Ehrlichia ruminantium"
```

```
isolate="Lemcot3"
/db_xref="taxon:779"
991..1839
/notes="ORF2"
/codon_start=1
/transl_table=11
/product="major antigenic protein 1 like protein"
/protein_id="AAD26349.1"
/db_xref="GI:4589104"
/transl_table="MNYKKILVRSALISLMSFLPYQSPAEVPVSNNTGNENAKGFFYI
SAKNPSPHPRKFSAEETPVYKDSPTKVKVGLKKEGSIKYSDFRTDTSFEGQNN
VLFSGSISYIMDPREVEAAYQKFNKPANETDTSYIKYKFSRAAMADKKY
WLNNGVTSSLMFNACDYDTAEGVFIPIVACAGIGADLISI FDDNLKFPAYOGKIG
ISVPIETSEISAFIGGYHGVIGNKYNKVPKLPVLTLDAPQSTASVTLDAFYEGEL
GVRFV"
2213..3058
/gene="MAP1"
2213..3058
/gene="MAP1"
/codon_start=1
/transl_table=11
/product="major antigenic protein 1"
/protein_id="AAD26350.1"
/db_xref="GI:4589105"
/transl_table="MNYKKIFITSLISLVSFLPQVFSDAIQBDSFVSGSVISAKY
MPTASHFKMSIKEDSRDQKVPGLKQDQVKTSSNTIFTEKDYSPKYNNTPLGF
AAGLYSMNGPRIEFSEIYETFDVKNPGYKNDAMHYCALDTSATSGGAAASVNV
KNEMLTDSLNLNACDYIMLDGMPVSPVCAGITGDLVSVINSPKLSYOGKUISY
SINPEASIFITGGHPRHVGNEFKDITTSKIFNTSGTGGTGPASATLDVCHFGIEIG
GRFV"
BASE COUNT 1292 a 449 c 478 g 1319 t
ORIGIN
Query Match 62.9%; Score 528.2; DB 1; Length 3538;
Best Local Similarity 78.2%; Pred. No. 8.7e-85;
Matches 664; Conservative 0; Mismatches 173; Indels 12; Gaps 2;
Qy 1 ATGAATATAGAAATCTAGTAGAGCGGTTAATCTCATTAATGCTCAATCTTACCA 60
Db 991 ATGAATATAAAAAATCTCGTAAGAAGTCGGTTAATCTCGTTAATGCTCAATCTTACCA 1050
Qy 61 TATCAGTCTTTTGCAGATCTCTGTAGGTTCAA-----GAACTAATGATAACAAGAA 111
Db 1051 TATCAGTCTTTTGCAGAACCTGTAAGTTCAATTAACATTTGGAATGAAATGCTAAAGAA 1110
Qy 112 GGTCTTACATTAGTGCAAGTACAATCCAAAGTATATCACATTTTAGAAAATTTCTCTGCT 171
Db 1111 GGGTTCTACATAAGTCAAAATACAAACCAAGCATACCCACACTTCAGAAAAATTTTCTGCT 1170
Qy 172 GAAGAAATCTCTATTATTAAGAACAAATTTCTCTCACTAAAAAGTTTTCGGACTAAAGAA 231
Db 1171 GAGGAAATCTCTGTATACGGTAAAGACTCTCCAACTAAAAAGGTATTTGGGTAAAAAAG 1230
Qy 232 GATGGTGATATAACAAAAAGACGATTTTACAAGAGTAGCTCCAGGCAATTGATTTTCAA 291
Db 1231 GAGGTTCTATAACAAATACAGTGAATTTCTACTAGAACAGATATATCGTTTGAGGGCCAA 1290
Qy 292 AATAACTTAATACAGGATTTTTCAGGAAGTATTTGGTTACTCTCTATGACGCGACCAAGATA 351
Db 1291 AATAATTTTATCTCAGGTTTCTCAGGAAGCATAGGTTATATCATGATGGACCAAGATA 1350
Qy 352 GAATCTGAAGCTGCATATCAACAATTTAATCCAAAAACACCGATAACAATGATCTGAT 411
Db 1351 GAGATTGAAGCTGCATACCAAAAATTCACCCAAAAATCCAGCT---AATGAAACTGAT 1407
Qy 412 AATGGTGAATACTATAAACAATTTTGCATTTATCTCGTAAAGATGCAATGGAAGATCAGCAA 471
Db 1408 ACTAGTGATTACTATAAACAATGATGATTTATCTCGTGAGAAGCTATGCGCAGATAAAAAA 1467
Qy 472 TATGTAGTACTTAAAAATGACGCGCATAACTTTTATGTCATTGATGCTGTTAATCTGCTAT 531
Db 1468 TATGTTGCTACTACAAATTAATGGAGTAACATTTTTCATCATTAATGTTTCAATGCTTCTAT 1527
```

```
QY 532 GACATTACAGCTGAGGAGTATCTTTCTGACCATATGATGCGAGGTATAGGAGCAGAT 591
Db 1528 GACATTACAGCAGAGGAGTACCTTTCTTCATATGATGATGCTGTTATTTGGTGCAGAT 1587

QY 592 CTTATCACTATTTTAAAGACTCAATCTAAAATTTGCTTACCAAGGAAAAATAGGTATT 651
Db 1588 CTTATATCTATTTGATGATATAAATTTAAATTTGCTTACCAAGGTAAATTTGGTATT 1647

QY 652 AGTTACCTATACACCAAGAAGTCTCTGCATTTTATTTGGTGGATACTACCATGGCGTTATT 711
Db 1648 AGTTATCAATTTACTCTCTGAAATTTCTGCAATTTATTTGGCGGATATTACCATGGAGTAATA 1707

QY 712 GGTATAAATTTGAGAAGATACCTGTATATACTCTCTGTAGTATTAATGATGCTCTCAAA 771
Db 1708 GGTAAACAGTATAACAAGTACCTGTAAAAGCTTCTGTAACTTTTAAACAGATGCTCTCTCAA 1767

QY 772 ACCACATCTGCTTCAGTAACTCTTGACGTTGGATCTTTTGGCGGAGAAATTTGGAAATGAGG 831
Db 1768 AGCACTTCGCTTCAGTAACTCTTGACGCTGGATATTTTGGTGTGAACCTTGGAGTAAGG 1827

QY 832 TTCACCTTC 840
Db 1828 TTTACTTTC 1836

RESULT 8
AF125279 LOCUS 3541 bp DNA linear BCT 20-APR-1999
DEFINITION Cowdria ruminantium isolate Antigua major antigenic protein 1 like
protein and major antigenic protein 1 (WAP1) genes, complete cds.
ACCESSION AF125279
VERSION AF125279.1 GI:4589109
KEYWORDS
SOURCE
ORGANISM Ehrlichia ruminantium.
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
1 (bases 1 to 3541)
Sulsona, C.R., Mahan, S.M. and Barbet, A.F.
The map1 Gene of Cowdria ruminantium Is a Member of a Multigene
Family Containing Both Conserved and Variable Genes
Biochem. Biophys. Res. Commun. 257 (2), 300-305 (1999)
2 (bases 1 to 3541)
Sulsona, C.R., Mahan, S.M. and Barbet, A.F.
Direct Submission
Submitted (02-FEB-1999) College of Veterinary Medicine Department
of Pathobiology, University of Florida, P.O. Box 110880,
Gainesville, FL 32610, USA
FEATURES
Location/Qualifiers
1..3541
/organism="Ehrlichia ruminantium"
/isolate="Antigua"
/db_xref="taxon:779"
989..1837
/note="ORF2"
/codon_start=1
/transl_table=11
/product="major antigenic protein 1 like protein"
/protein_id="AAD26353.1"
/db_xref="GI:4589110"
/translation="MNKKILVRSAISLMSFLPQSFAPVSVNNIGNENAKGFGYI
SAKNPSPHPRKSAEETPVYKDSPTKKVFLKKGSTIKYSDFTRTDIDISFEGQNN
FISGFSIGYIMDGPRIEAYQKFNPKNPANETDTSYKHVGLSRAEMADKKY
VLTNGVTFSLSMFNACDITAEGVVPIPVACAGIGADLISIFDDINLKFAYQKIG
ISYPTPEISAFIGGYHGVIGNKYNKVPVKLPVLTIDAPQSTASVTLDAAGYGGEL
GVRTFP"
2210..3055
/gene="WAP1"
2210..3055
/gene="WAP1"
/codon_start=1
/transl_table=11
/product="major antigenic protein 1"
gene
CDS
831
CDS
1825
832 TTCACCTTC 840
|||
|||
```

```
/protein_id="AAD26354.1"
/db_xref="GI:4589111"
/translation="MNCKKIFITSTLISLVSFLPGVSFSDVIQDSSPVGSVYISAKY
MPTASHFGKMKIKEDSKDVKVFGKLDGDKVKTSSNTITFKEDKYSFKYNNPFLGF
APTAIGYMGNGPRIEFELSYETFDVKNPGNVKNDAMVYCALDTATSSGAAASTVMV
KNENLTDLSIMLNACVDIMLDGHPVSPVYVAGIGTDLVSVLNSNPKLSYQKLGISY
SINPEASIFGHHFHRVIGNEFKDITTSKIFNTSNTGATPGFASAILDUDVCHFIEIG
GRPVF"
BASE COUNT 1282 a 451 c 479 g 1329 t
ORIGIN
Query Match 62.9%; Score 528.2; DB 1; Length 3541;
Best Local Similarity 78.2%; Pred. No. 8.7e-85;
Matches 664; Conservative 0; Mismatches 173; Indels 12; Gaps 2;
QY 1 ATGAATTTATAGAAAAATTTCTAGTAAGAGGCGGTTAATCTCATTAATGTCTCAATCTTACCA 60
Db 989 ATGAATTTATAGAAAAATTTCTGTAAGAGGCGGTTAATCTCATTAATGTCTCAATCTTACCA 1048
QY 61 TATCAGTCTTTTGGAGATCTCTGTAGTTCAA-----GAACATAATGATAACAAGAA 111
Db 1049 TATCAGTCTTTTGGAGAACCTGTAAAGTTCAAATAACATTTGAAATGAAATGCTAAAGAA 1108
QY 112 GGCTTCTACATTAGTGCAGAGTACAATCCAGTATATACACATTTAGAAAAATTTCTCTGCT 171
Db 1109 GGGTCTTACATAAGTGCAGAAATACACCAAGCATACCACTTTCAGAAAAATTTTCTGCT 1168
QY 172 GAAGAACTCTCTAATTAATGAACAAATTTCTCTCACTAAAAAAGTTTTCGGACTAAAGAA 231
Db 1169 GAGGAACTCTCTGTATACGTTAAGACTCTCCAACTAAAAAGGTATTTGGGTTAAAAAAG 1228
QY 232 GATGCTGATATAACAAAAAAGACGATTTTACAAGAGTAGCTCCAGGCAATTTGATTTTCAA 291
Db 1229 GAGGTTCTATAACAAAAATACAGTGAATTCACATAGAACAGATATATCGTTTGGAGGCCAA 1288
QY 292 AATTAACCTAATATCAGGATTTTTCAGGAAGTATTTGGTTACTCTATGGACGACCAAGATA 351
Db 1289 AATAAATTTTATCTCAGGTTTCTCAGGAAGCATAGGTATATATCATGGATGGACCAAGATA 1348
QY 352 GAACCTGAAGCTGATATCAACAAATTTAATCCAAAAACACCGATAACAATGATCTCAT 411
Db 1349 GAGATTGAGCTGATATACCAAAATTTCAACCCAAAAAATTCAGCT--AATGAACCTGAT 1405
QY 412 AATGGTGAATATAAACAATTTTTCATTAATCTCGTAAAGATGCAATGGAAGATCAGCAA 471
Db 1406 ACTAGTGATTACTATAAACAATATGGATTATCTCGTCAGAAAGCTATGGCAGATAAAAAA 1465
QY 472 TATGTAGTACTTAAAAATGACGCAATACTTTTATGTCATTTGATGTTTAACTACTTCTAT 531
Db 1466 TATGTGTACTTACAAATAATGGAGTAACCTTTTTCATCATTAATGTTCAATGCTTGTAT 1525
QY 532 GACATTACAGCTGAAGGAGTATCTTTCTGACCATATGATGCGAGGTATAGGAGCAGAT 591
Db 1526 GACATTACAGCAGAGGAGTACCTTTTCAATCATGATGATGCTGTTATTTGGTGCAGAT 1585
QY 592 CTTATCACTATTTTAAAGACCTCAATCTAAAATTTGCTTACCAAGGAAAAATAGGTATT 651
Db 1586 CTTATATCTATATTGATGATATAAATTTAAATTTGCTTACCAAGGTAAATTTGGTATT 1645
QY 652 AGTTACCTATACACCAAGAGTCTCTGCATTTTATTTGGTGGATACTACCATGGCGTTATT 711
Db 1646 AGTTATCCAATTAATCTCTGAAATTTCTGCATTTTATTTGGCGGATATTACCATGGAGTAATA 1705
QY 712 GGTATAAATTTGAGAAGATACCTGTAACTCTCTGTAGTATTAATGATGCTCTCTCAA 771
Db 1706 GGTAAACAGTATAACAAGTACCTGTAAAGCTTCTGTAACTTTTAAACAGATGCTCTCTCAA 1765
QY 772 ACCACATCTGCTTCAGTAACTCTTGACGTTGGATCTTTTGGCGGAGAAATTTGGAAATGAGG 831
Db 1766 AGCACTTCGCTTCAGTAACTCTTGACGCTGGATATTTTGGTGTGAACCTTGGAGTAAGG 1825
QY 832 TTCACCTTC 840
|||
|||
```

```

Db      1826 TTTACTTTC 1834

RESULT 9
AF125275
LOCUS
DEFINITION
  3551 bp DNA linear BCT 20-APR-1999
  Cowdria ruminantium isolate Crystal Springs major antigenic protein
  1 like protein and major antigenic protein 1 (MAP1) genes, complete
  cds.
ACCESSION
  AF125275
VERSION
  AF125275.1 GI:4589097
KEYWORDS
  Ehrlichia ruminantium.
SOURCE
  Ehrlichia ruminantium.
  Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
  Anaplasmataceae; Ehrlichia.
REFERENCE
  1 (bases 1 to 3551)
  Sulsona, C.R., Mahan, S.M. and Barbet, A.F.
  The map1 Gene of Cowdria ruminantium Is a Member of a Multigene
  Family Containing Both Conserved and Variable Genes
  Biochem. Biophys. Res. Commun. 257 (2), 300-305 (1999)
JOURNAL
  Submitted (02-FEB-1999) College of Veterinary Medicine Department
  of Pathobiology, University of Florida, P.O. Box 110880,
  Gainesville, FL 32610, USA
FEATURES
  source
    Location/Qualifiers
      1. 3551
        /organism="Ehrlichia ruminantium"
        /isolate="Crystal Springs"
        /db_xref="taxon:779"
        991..1839
        /note="ORF2"
        /codon_start=1
        /transl_table=11
        /product="major antigenic protein 1 like protein"
        /protein_id="AAD26345.1"
        /db_xref="GI:4589098"
        /translation="MNYKKILVRSALISLMSFLPYQSPAEVPVSSNNIGNAKEGFYI
        SAKNPISIPHRKFSAEPTVYKDSPTKVLKKEGSIKYSDFTDLSFEGQNN
        FISGFSIGSIVIMDPRVEIEAAVOKFNPKNANETDTSVYKHVGLSRAETMTDKKY
        VLVNNGVTFSSLMFNACVDITAEQVPEIPVACAGIGADLJLIFDDINLKFAVYOKGIG
        ISYPTTPEISAFIGGYHGVIGNKINKVPVKLPVLTLDAPQSTASVTLDAVYGGEL
        GVRTTF"
      2199..3062
        /gene="MAP1"
      2199..3062
        /gene="MAP1"
        /codon_start=1
        /transl_table=11
        /product="major antigenic protein 1"
        /protein_id="AAD26346.1"
        /db_xref="GI:4589099"
        /translation="MNCCKIFITSLISLVSLPGVPSDVIQEDSNPAGSVVISAKY
        NPTASHFGKMSIKEDSKNTQVFLKKDWGDKVTPSDSSNTNFIIFTEKDYSPRYENN
        PFLGAGAGIYSMGPRIEFVSYETFDVKNPGNKKNNHMYCALDTAAQNSNGAG
        LITSVMVKNELNLSIMLNACVDIMLDGIPVYVCAGIGTDLVLSVINATNPKLISYQ
        KGLGISYINSEASIGFICGHFHRVIGNEFKDIATLKIFTSKTSINPGFASATLDVCH
        FGIIGGRFV"
BASE COUNT 1308 a 477 c 464 g 1302 t
ORIGIN

Query Match 62.9%; Score 528.2; DB 1; Length 3551;
Best Local Similarity 78.2%; Pred. No. 8.7e-85;
Matches 664; Conservative 0; Mismatches 173; Indels 12; Gaps 2;

Qy      1 ATGAATTATACAAAATTTCTAGTAGAAGCGGTTAAATCTCAATTAATGTCATCTTACCA 60
        |||||
Db      991 ATGAATTATACAAAATTTCTGTAAGAAGTGGGTTAAATCTCAATTAATGTCATCTTACCA 1050
        |||||
Qy      61 TATCAGTCTTTTGCAGATCCTCTAGGTTCAA-----GAACATAATGATAACAAAGAA 111
        |||||

```

1051 TATCAGTCTTTTGCAGAACCTGTAAAGTTCAAAATAACATTGGAAATGAAATGCTAAAGAA 1110

112 GGCTTTACATTAAGTCAAAAGTACAAATCAAGTATATCAACATTTAGAAAATTTCTGCT 171

1111 GGGTTTACATAAGTCAAAATATCAACCAAGCATACCACTTCAGAAAAATTTTCTGCT 1170

172 GAAGAAATCTCTATTAAATGGAACAAATTTCTCTACTAAAAAGTTTTCGGACTAAAGAAA 231

1171 GAGGAAATCTCTGTATACGGTAAAGACTCTCCAACTAAAGAGTTATTTGGGTTAAAAAAG 1230

232 GATGCTGATATAACAAAAAGACGATTTTACAAGAGTAGCTCCAGGCAATTCATTTTCAA 291

1231 GAGGTTCTATAACAAATACAGTGAATTTCACTAGACAGATATATCGTTTGGAGGCCAA 1290

292 AATACTTAATATCAGGATTTTCAAGAGTATTTGGTACTCTATGAGCGGACCAAGAATA 351

1291 AATAATTTTATCTCAGGTTTCTCAGGAAGCATAGGTTATATCATGGATGGACCAAGATA 1350

352 GAATCTGAAGCTGCATATCAACAAATTTAATCCAAAAACACCGATAACAATGATCTGAT 411

1351 GAGATTGAAGCTGCATACCAAAATTTCAACCCAAAAATCCAGCT---AATGAACCTGAT 1407

412 AATGCTGAATACTATAAACATTTTTCATTTCTCGTAAAGATGCAATGAAGAGATCAGCAA 471

1408 ACTAGTACTATAAACACTATGGATTATCTCGTCAGAAACCATGACAGATAAAAAA 1467

472 TATGTAGTACTTAAAAATGACGGCATAACTTTTATGTCATTTGATGGTTTAATCTGCTAT 531

1468 TATGTTGTTACTTACAAATTAATGGAGTAACTTTTTCATCATCATTAATTAATGCTTGTAT 1527

532 GACATTACAGCTGAAGGAGTATCTTCTGCATCATATGCTGCGAGGTATAGGAGCAGAT 591

1528 GACATTACAGCAGGAGGAGTACCTTTTCATTCATGATGCTGCTGTTTGGTGCAGAT 1587

592 CTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGAAAAATAGGTATT 651

1588 CTTATATCTATTTGATGATATAAATTTAAAAATTTGCTTACCAAGGTAAAAATTTGGTATT 1647

652 AGTTACCTTATCACACAGAAAGTCTCTGCATTTATTTGGTGGATATACCATGGGTTATT 711

1648 AGTTATCAATTAATCTCTGAAATTTCTGCATTTATTTGGCGGATATACCATGGAGTAATA 1707

712 GGTATAAATTTGAGAAGATACCTGTAATACTCCTGTAGTATTAATATGATGCTCCTCAA 771

1708 GGTAAACAGTATAACAAGTACCTGTAAGCTTCTGTAACTTTAAACAGATGCTCCTCAA 1767

772 ACCATCTGCTTCAGTAACTCTTTCAGCTTGGATACATTTTGGCGGAGAAATGGAATGAGG 831

1768 AGCACTTCGCTTCAGTAACTCTTTCAGCTGATATTTTGGTGTGAATTTGGAGTAAAGG 1827

832 TTCACCTTC 840

1828 TTTACTTTC 1836

RESULT 10

AF125278

LOCUS

DEFINITION

3572 bp DNA linear BCT 20-APR-1999

Cowdria ruminantium isolate UmBanein major antigenic protein 1 like

protein and major antigenic protein 1 (MAP1) genes, complete cds.

ACCESSION

AF125278

VERSION

AF125278.1 GI:4589106

KEYWORDS

Ehrlichia ruminantium.

SOURCE

Ehrlichia ruminantium

ORGANISM

Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

Anaplasmataceae; Ehrlichia.

REFERENCE

1 (bases 1 to 3572)

Sulsona, C.R., Mahan, S.M. and Barbet, A.F.

The map1 Gene of Cowdria ruminantium Is a Member of a Multigene

Family Containing Both Conserved and Variable Genes

Biochem. Biophys. Res. Commun. 257 (2), 300-305 (1999)

JOURNAL

Submitted (02-FEB-1999) College of Veterinary Medicine Department

of Pathobiology, University of Florida, P.O. Box 110880,

Gainesville, FL 32610, USA

FEATURES

source

Location/Qualifiers

1. 3572

/organism="Ehrlichia ruminantium"

/isolate="Crystal Springs"

/db\_xref="taxon:779"

991..1839

/note="ORF2"

/codon\_start=1

/transl\_table=11

/product="major antigenic protein 1 like protein"

/protein\_id="AAD26345.1"

/db\_xref="GI:4589098"

/translation="MNYKKILVRSALISLMSFLPYQSPAEVPVSSNNIGNAKEGFYI

SAKNPISIPHRKFSAEPTVYKDSPTKVLKKEGSIKYSDFTDLSFEGQNN

FISGFSIGSIVIMDPRVEIEAAVOKFNPKNANETDTSVYKHVGLSRAETMTDKKY

VLVNNGVTFSSLMFNACVDITAEQVPEIPVACAGIGADLJLIFDDINLKFAVYOKGIG

ISYPTTPEISAFIGGYHGVIGNKINKVPVKLPVLTLDAPQSTASVTLDAVYGGEL

GVRTTF"

2199..3062

/gene="MAP1"

2199..3062

/gene="MAP1"

/codon\_start=1

/transl\_table=11

/product="major antigenic protein 1"

/protein\_id="AAD26346.1"

/db\_xref="GI:4589099"

/translation="MNCCKIFITSLISLVSLPGVPSDVIQEDSNPAGSVVISAKY

NPTASHFGKMSIKEDSKNTQVFLKKDWGDKVTPSDSSNTNFIIFTEKDYSPRYENN

PFLGAGAGIYSMGPRIEFVSYETFDVKNPGNKKNNHMYCALDTAAQNSNGAG

LITSVMVKNELNLSIMLNACVDIMLDGIPVYVCAGIGTDLVLSVINATNPKLISYQ

KGLGISYINSEASIGFICGHFHRVIGNEFKDIATLKIFTSKTSINPGFASATLDVCH

FGIIGGRFV"

BASE COUNT 1308 a 477 c 464 g 1302 t

ORIGIN

Query Match 62.9%; Score 528.2; DB 1; Length 3551;

Best Local Similarity 78.2%; Pred. No. 8.7e-85;

Matches 664; Conservative 0; Mismatches 173; Indels 12; Gaps 2;

Qy 1 ATGAATTATACAAAATTTCTAGTAGAAGCGGTTAAATCTCAATTAATGTCATCTTACCA 60

|||||

Db 991 ATGAATTATACAAAATTTCTGTAAGAAGTGGGTTAAATCTCAATTAATGTCATCTTACCA 1050

|||||

Qy 61 TATCAGTCTTTTGCAGATCCTCTAGGTTCAA-----GAACATAATGATAACAAAGAA 111

|||||

Qy	412	AATGGTGAATACATATATAAACATATTTTGGCATATATCTCGTAAAGATGCAATGGGAAGATACAGAA	471
Db	1406	ACTAGTGAATACATATATAAACACTATGGATTATCTCGTGCAGAAGCTATGGCAGATAAAAAA	1465
Qy	472	TATGTAGTACTTTAAAATGACGCATAACTTTTATGTCATTTATGTCATTTATGTTATGTTATGCTAT	531
Db	1466	TATGTTGTACTTACAAATATATGGAGTAACCTTTTTCATCATTAATGTTTAAATGCTTTGCTAT	1525
Qy	532	GACATTACAGCTGAGGAGATATCTTTTCGTACCATATATGTCATGTCAGGATATAGGAGCAGAT	591
Db	1526	GACATTACAGCAGAGAGTACCTTTTCATTCATATGTCATGTCGCTGGTATGGTGCAGAT	1585
Qy	592	CTTATCACATATTTTAAAGACCTCAATCTAAATAATTTGCTTTACCAAGAAATAAGGTATT	651
Db	1586	CTTATATCTATATTTGATGATATAAATTTTAAAAATTTTGCCTTACCAAGGTAATAATTTGGTATT	1645
Qy	652	AGTTACCTCTATCACACGAGAGTCTCTGCATTTATTTGTTGATACCTACCATGGCGTTATT	711
Db	1646	AGTTATCCAAATTAATCTCTGAAATTTCTGCAATTTATTTGGCGGATATACCATGGAGTAATA	1705
Qy	712	GGTAATAAATTTTCCAGAAAGATACCTGTAAATAACCTCTGTAGTATTTAAATGATGCTCTCTCAA	771
Db	1706	GGTAACAGATATAACAAGTACCTGTAAAGCTTCTGTAACTTTTACAGATGCTCTCTCAA	1765
Qy	772	ACCACATCTGCTTCAGTAACTCTTTGACGTTGGATACCTTTGGCGAGAAATTTGGAATGAGG	831
Db	1766	AGCACTTCGCTTCAGTAACTCTTTGACGCTGGATATTTTGGTGGTGAATTTGGAGTAAGG	1825
Qy	832	TTACAGCTTC 840	
Db	1826	TTTACTTTC 1834	
RESULT 11			
AF319940			
LOCUS			
DEFINITION			
AF319940			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
Medicine, Yalelaan 1, Utrecht 3584CL, The Netherlands			
FEATURES			
Source			
gene			
CDS			

```
/db_xref="GI:17224942"
/translation="KLAQVQKLGISYPMSSNIKLPADQYHYKVTNDKFNKLVKVIHVD
LYNPLVTSATATLNIEYFGGEIGIRFGLK"
1613..>2362
/gene="map1-1"
1613..>2362
/gene="map1-1"
/codon_start=1
/transl_table=11
/product="major antigenic protein 1-1"
/protein_id="RAL37180.1"
/db_xref="GI:17224943"
/translation="MNYKKILVRSALISLMSFLPVQSPAEVPVSSNNIGNENAKSGFYI
SAKNPSIPHRKSAETPYGKDSPTKVFGLKKDGSITKYSDPTFDISLSPGQSN
VLTSGSIGIMDGPVEIEAAYOKFNPKNPANETDTSYIKHGLSRAETMDKKY
FVLNNGVTFSLFMNACYDITAEGVPIYPACAGIGADLISIFDDINLKPAYQKIG
ISYPTIPRISAFIGYHGVGNKYKIPVK"
BASE COUNT 939 a 295 c 263 g 863 t 2 others
ORIGIN
Query Match 54.5%; Score 457.6; DB 1; Length 2362;
Best Local Similarity 77.8%; Pred. No. 3.7e-72;
Matches 582; Conservative 0; Mismatches 154; Indels 12; Gaps 2;
QY 1 ATGAATTATAAGAAATCTAGTAGAAGCGGTTAATCTCAATTAATGTCATCTTACCA 60
DB 1613 ATGAATTATAAAAAATCTCGTAGAAGTGGTTAATCTCAATTAATGTCATTTTACCA 1672
QY 61 TATCAGTCTTTTGCAGATCCTGTAGGTCAA-----GAACTAATGATAACAAGAA 111
DB 1673 TATCAGTCTTTTGCAGAACCTGTAAGTTCAATAACATTTGGAATGAAATGCTAAAGAA 1732
QY 112 GCCTCTACATTAGTGCAAGTACAAATCAAGTATATACACATTTAGAAAATCTCTGCT 171
DB 1733 GGGTCTTATAGTGCAGAAATACAAACCAAGCATACCAATTTTCAGAAAAGTTTCTGCT 1792
QY 172 GAAGAAACCTCCTATTAAATGGAACAAATCTCTCACTTAAAGAGTTTTCGGACTAAGAA 231
DB 1793 GAGGAACCTCCTGTATAGCGTAAAGACTCTCAACTTAAAGAGTATTTGGATTGAAAAG 1852
QY 232 GATGGTGATATAACAAAAAGACGATTTTACAAGAGTAGTCTCCAGGCGATTGATTTTCAA 291
DB 1853 GACGGTCTATAACAAATACAGTGATTTCTAGTACACAGATATATCGTTTGAGGCGCAA 1912
QY 292 AATAACTTAATACAGGATTTTCAGAGATATTTGTTACTCTATGCGAGCGACCAAGATA 351
DB 1913 AGTAATTTTATCTCAGGTTTCTCAGGAAGCATAGGTATATCATGCGGACCAAGAGTA 1972
QY 352 GAACTTGAAGCTGCATATCAACAAATTTAATCCAAAAACACCGATAACAAATGATCTGAT 411
DB 1973 GAATTTGAAGCTGCATATCCAAAATTTCAACCCAAAATATCCAGCT---AATGAACTGAT 2029
QY 412 AATGGTGAATACTATAAAACATTTTGCAATTTATCTCGTAAAGATGCAATGGAAGATCAGCAA 471
DB 2030 ACTAGTGATTACTATAAACACATATGATTTATCTCGTGCAAGAACCATGACAGATAAAAA 2089
QY 472 TATGTAGTACTTAAAAATGACGGCATAACTTTTATGCTATGATGGTTAATATCTGCTAT 531
DB 2090 TATGTGTACTTTACAAATTAATGGAGTAACTTTTTCATCAATTAATGTTTAAACGCTTGCTAT 2149
QY 532 GACATTACAGCTGAAGGAGTATCTTTTCGPACCATATGCAATGCGAGGTATAGGACAGAT 591
DB 2150 GATATTACAGCAGAGGAGTACCTTTTCATTCATATGCATGCGCTGGTATTTGGTCAGAT 2209
QY 592 CTTATCACTATTTTAAAGACCTCAATCTTAAATTTTGCTTACCAAGGAAAAATAGGTATT 651
DB 2210 CTTATATCTATATTTGATGATATAAATTTTAAATTTTGCTTACCAAGGTAATAATGCTATT 2269
QY 652 AGTTACCTATACACAGCAAGTCTCTGCATTTTATTTGGTGATATACCATGGCGCTATT 711
DB 2270 AGTTATCAACATCACTCTCTGCAATTTTCTGCATTTATTTGGCGGATATATACCATGGAGTAATA 2329
QY 712 GGTATAAAATTTGAGAAGATACCTCTGTA 739
```

```
DB 2330 GGTATAAGTATATAATAAATACCTGTAA 2357
|||||
RESULT 12
AF082745 Ehrlichia canis strain Louisiana 834 bp DNA linear BCT 26-JUL-2000
LOCUS gene, partial cds.
DEFINITION Ehrlichia canis strain Louisiana 834 bp DNA linear BCT 26-JUL-2000
ACCESSION AF082745
VERSION AF082745.1 GI:3769524
KEYWORDS Ehrlichia canis.
SOURCE Ehrlichia canis.
ORGANISM Ehrlichia canis
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE 1 (bases 1 to 834)
AUTHORS McBride, J.W., Yu, X. and Walker, D.H.
TITLE Molecular cloning of the gene for a conserved major immunoreactive
28-kilodalton protein of Ehrlichia canis: a potential
serodiagnostic antigen
JOURNAL Clin. Diagn. Lab. Immunol. 6 (3), 392-399 (1999)
MEDLINE 99242757
PUBMED 10225842
REFERENCE 2 (bases 1 to 834)
AUTHORS McBride, J.W., Yu, X.J. and Walker, D.H.
TITLE Direct Submission
JOURNAL Submitted (07-AUG-1998) Pathology, University of Texas Medical
Branch, 301 University Blvd., Galveston, TX 77555-0609, USA
FEATURES
source
1..834
/organism="Ehrlichia canis"
/strain="Louisiana"
/db_xref="taxon:944"
1..>834
/note="p28"
/codon_start=1
/transl_table=11
/product="28 kDa outer membrane protein"
/protein_id="AAC64551.1"
/db_xref="GI:3769525"
/transl_table="MNCCKILITLALISLMYSIPSFSDTIQDNMGNFYIISQYV
PSVSHFGSFSKESKSTGVFLGKHWGDSPILNKHADFTVNYSPRYNNPFLGK
AGATGYSFGSGPREFEISYEAFOVKSPNINQNDARYCALSHSTSAAMEADKVFLLK
NEGILSLAINACYDIINDKVPVSPYICAGIGTDLISMFEATSPKISYQKLGISYS
INPSTVFIQGHFHRIIGNEFRDIPALVPSNSTITSGPQFATVLNVCHFLELGGFR
NF"
BASE COUNT 277 a 147 c 151 g 259 t
ORIGIN
Query Match 28.4%; Score 238.4; DB 1; Length 834;
Best Local Similarity 57.7%; Pred. No. 5.4e-33;
Matches 486; Conservative 0; Mismatches 346; Indels 10; Gaps 3;
QY 1 ATGAATTATAAGAAAATTTCTAGTAAGAACGCGGTTAATCTCAATTAATGTCATCTTACCA 60
DB 1 ATGAATTCGCAAAAAAATTTCTTATACAACTGCATTAATATATCATTAATGATCTTATTTCA 60
QY 61 TATCAGTCTTTTGCAGATCCTGTAGGTTTCAAGAACTAATGATAACAAGAGGCTTCTTAC 120
DB 61 AGCATATCTTTTCTGATACCTATACAGATGCTAACATGGTGGGTGTAA-----CTTCTAT 114
QY 121 ATTAGTGCAAGTACAAATCCAACTATATACACATTTAGAAAATTTCTGCTGAAGA--AA 178
DB 115 ATTAGTGCAAGTATGATCAAGAGTGTCTCAATTTTGGTAGCTTCTCAGCTTAAAGAGAA 174
QY 179 CTCCTTATTAATGGAACAAATTTCTCTCACTAAAAAGTTTTCGGACTAAAGAAAGATGGTG 238
DB 175 AGCAATCAACTGTGGAGTTTGGATTAAAAATGATGTTGGATGGAGTCCCACTACT- 233
QY 239 ATATAACAAAAAGACGATTTTACAAGAGTAGCTCCAGGAGTATGATTTTCAAAATAACT 298
DB 234 -TAAGATAAACACGCTGACTTTACTGTTCCTCAAACTATTCTGTTTCCAGATACGAGAACATC 292
```

```
QY 299 TAATATCAGGATTTTCAGGAAGTATTGGTTACTCTATCGACGACCAAGAAATAGAACTTG 358
Db 293 CATTTCTAGGTTTTCAGGAGCTATCGGTTACTCAATGGGTGGCCCAAGAAATAGAAATCG 352
QY 359 AAGCTGCATATCAACAATTTAATCAAAAAACACCGATAACAAGTACTGATTAATGGTG 418
Db 353 AAATATCTTATGAAGCATTCGACGTAAAGATCTCTAATATCAATATCAAAATGACGGC 412
QY 419 AATACATAAACAATTTGTCATTTCTCGTAAGATGCAATGCAAGATCAGCAATATGTAG 478
Db 413 ACAGTACTCGCTCTATCTCATCACATCGGCAGCCATGGAAGCTGATAAATTTGTCT 472
QY 479 TACTTAAATGACGGCAATCTTTATGTCAATGATGGTTAATATCTGCTATGACATTA 538
Db 473 TCTTAAAAACGAAGGTTAATGACATATCACTTGCAATAAATGCAATGTTATGATATA 532
QY 539 CAGCTGAAGGAGTATCTTTCTGACCATATGATGTCAGGTATAGGAGCAGATCTTATCA 598
Db 533 TAAATGACAAAGTACCTGTTCTCTTATATATGCGCAGGTATGGTACTGATTTGATTT 592
QY 599 CTATTTTAAAGACCTCAATCTAAATTTGCTTACCAGGAAATAGGTATTAGTTACC 658
Db 593 CTATGTTGAAGTACAGTCTTAAATTTCTTACCAGGAAATAGGTATTAGTTACT 652
QY 659 CTATCACACGAGTCTCTGATTTATGTTGATGACTACATGCGGCTTATGGTAAATA 718
Db 653 CTATTAATCCGAAACCTCTGTTTCTATGCGGTGGCATTTTCACAGGATCATAGTAATG 712
QY 719 AATTGAGACATCTGTAATACCTCTGATGATTAATTAATGATGCTCTCAACCCACNT 778
Db 713 AGTTAGAGATTTCTCTGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 772
QY 779 CTGCTTCAGTAACCTTTGACGCTTGGATCTTTGGCGGAGAAATGGAATGAGGTTACCT 838
Db 773 TTGCAACAGTAACACTAAATGCTGTCATCTTGGTTAGACTTTGGAGGAGATTTAACT 832
QY 839 TC 840
Db 833 TC 834
```

```
RESULT 13
AF082746 834 bp DNA linear BCT 26-JUL-2000
LOCUS Ehrlichia canis strain Oklahoma 28 kDa outer membrane protein gene,
DEFINITION partial cds.
ACCESSION AF082746
VERSION AF082746.1 GI:3769526
KEYWORDS
SOURCE
ORGANISM Ehrlichia canis.
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE 1 (bases 1 to 834)
AUTHORS McBride, J.W., Yu, X. and Walker, D.H.
TITLE Molecular cloning of the gene for a conserved major immunoreactive
28-kilodalton protein of Ehrlichia canis: a potential
serodiagnostic antigen
JOURNAL Clin. Diagn. Lab. Immunol. 6 (3), 392-399 (1999)
MEDLINE 99242757
PUBMED 10225842
REFERENCE 2 (bases 1 to 834)
AUTHORS McBride, J.W., Yu, X.J. and Walker, D.H.
TITLE Direct Submission
JOURNAL Submitted (07-AUG-1998) Pathology, University of Texas Medical
Branch, 301 University Blvd., Galveston, TX 77555-0609, USA
FEATURES
source
1. .834
/organism="Ehrlichia canis"
/strain="Oklahoma"
/db_xref="taxon:944"
1. .834
CDS
```

```
/note="p28"
/codon_start=1
/transl_table=11
/product="28 kDa outer membrane protein"
/protein_id="AAC64552.1"
/db_xref="GI:3769527"
/translation="MNCCKILITLALISLMYSPISFSDTIQDNMGNGNPIYISGVV
PSVHFGSFAKESKSTGVFLKHDWDGSPILKNHADFVFNYSFRVNNPFLGP
AIGAYSMGPRIFEFISYFAFVSPENYNDHAYRICALSHSTSAAMEADKVFLLK
NEGILISLAINACYDIINDKVPVSPYICAGIGTDLISMPFATSPKTSYOCKLGISYS
INPETSVPFIGGHRHRIIGNEFRDIPAIVPSNSTTISGPQFATVTLNVCHFGLGGRF
NF"
BASE COUNT 277 a 147 c 151 g 259 t
ORIGIN
Query Match 28.4%; Score 238.4; DB 1; Length 834;
Best Local Similarity 57.7%; Pred. No. 5.4e-33;
Matches 486; Conservative 0; Mismatches 346; Indels 10; Gaps 3;
QY 1 ATGAATTTAAGAAAATTTCTAGTAAGAGCGGTTAATCTCATTAATGTCTTATACCA 60
Db 1 ATGAATTTGCAAAAAATTTCTTATAACAACCTGCATTAATATCATTAATGTCTTATCCA 60
QY 61 TATCAGTCTTTTGCAGATCTCTAGGTTCAAGAACTAATGATAACAAGAAAGGCTTCTAC 120
Db 61 AGCATATCTTTTCTGATCTACTATACAAGATGGTAACATGGGTGGTAA-----CTTCTAT 114
QY 121 ATTAGTCAAAAGTACAACTCAAGTATATCACCTTTTAGAAAAATTTCTGCTGAAGA--AA 178
Db 115 ATTAGTGAAGATGATGTACCAAGTCTCTCATTTTGGTAGCTTCTCAGCTAAAGAGAA 174
QY 179 CTCTATTAAATGGAAACAAATTTCTCTACTAAAAAAGTTTTCGGACTAAAGAAAGATGGTG 238
Db 175 AGCAAACTCAACTGTGGAGTTTGGATTTAAACATGATTTGGGATGGAAGTCCAACTACT- 233
QY 239 ATATAACAAAAAGACGATTTTACAAGATAGTCTCCAGGATTTGATTTTCAAAATAACT 298
Db 234 -TAAGAAATAACACGCTGACTTTTACTGTTCCAAACTATTCGTTCCAGATACGAGAACATC 292
QY 299 TAATATCAGGATTTTTCAGGAAGTATTGGTTACTCTATGGACGGACCAAGAAATAGAACTTG 358
Db 293 CATTTCTAGGTTTTCAGGAGCTATCGGTTACTCAATGGGTGGCCCAAGAAATAGAAATCG 352
QY 359 AAGCTGCATATCAACAATTTAATCCAAAAACACCGATAACAATGATGATGATGATGATG 418
Db 353 AAATATCTTATGAAGCATTCGACGTAAAGATCTCTAATAATCAATTTATCAAAATGACGGC 412
QY 419 AATACTATAACAATTTTGCATTTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTAG 478
Db 413 ACAGGTAATCGCTCTATCTCATCACATCGGCAGCCATGGAAGCTGATAAATTTGTCT 472
QY 479 TACTTAAAAATGACGGCATAAATTTTATGTCATTTGATGGTTAATATCTGCTATGACATTA 538
Db 473 TCTTAAAAACGAAGGTTAATGACATATCACTTGCATTAATGCAATGATGATGATGATAA 532
QY 539 CAGCTGAAGGAGTATCTTTCTGACCATATGTCAGGTATAGGAGTATAGGAGTATGATCA 598
Db 533 TAAATGACAAAGTACCTGTTTCTCTTATATATGCGCAGGTATTTGGTACTGATTTGATTT 592
QY 599 CTATTTTAAAGACCTCAATCTTAAATTTGCTTACCAGGAAATAGGTATTAGTTACC 658
Db 593 CTATGTTTGAAGCTTACAAGTCTTAAATTTCTTACCAGGAAATAGGTATTAGTTACT 652
QY 659 CTATCACACGAGGATCTCTGCAATTTATGCTGATGATGATGATGATGATGATGATGATG 718
Db 653 CTATTAATCCGAAACCTCTGTTTCTATGCGGTGGCATTTTCCACAGGATCATAGTAATG 712
QY 719 AATTGAGAAAGATACCTGTGTAATAACTCTCTGATGATTTAAATGATGCTCTCAAAACACAT 778
Db 713 AGTTAGAGATTTCTCTGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 772
QY 779 CTGCTTCAGTAACCTTTGACGCTTGGATCTTTGGCGGAGAAATGGAATGAGGTTACCT 838
Db 773 TTGCAACAGTAACACTAAATGCTGTCATCTTGGTTAGACTTTGGAGGAGATTTAACT 832
```

Db 773 TTGCAACAGTAACACATAATGTGTCTCACTTTGGTTAGAACTTGGAGGAGATTTAAC 832

Qy 839 TC 840

Db 833 TC 834

RESULT 14

AF082747

LOCUS

DEFINITION Ehrlichia canis strain Demon (North Carolina) 28 kDa outer membrane protein gene, partial cds.

ACCESSION AF082747

VERSION AF082747.1 GI:3769528

KEYWORDS

SOURCE

ORGANISM Ehrlichia canis.

Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Anaplasmataceae; Ehrlichia.

REFERENCE 1 (bases 1 to 834)

McBride, J.W., Yu, X.J. and Walker, D.H.

Molecular cloning of the gene for a conserved major immunoreactive 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic antigen

Clin. Diagn. Lab. Immunol. 6 (3), 392-399 (1999)

JOURNAL

MEDLINE

PUBMED 10225842

REFERENCE 2 (bases 1 to 834)

McBride, J.W., Yu, X.J. and Walker, D.H.

Direct Submission

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

BASE COUNT 277 a 147 c 151 g 259 t

ORIGIN

Query Match 28.4%; Score 238.4; DB 1; Length 834;

Best Local Similarity 57.7%; Pred. No. 5.4e-33;

Matches 486; Conservative 0; Mismatches 346; Indels 10; Gaps 3;

Qy 1 ATGAATTATAAGAAAATTTCTAGTAAAGACGCGTTAATCTCAATTAATGTCAATCTTACCA 60

Db 1 ATGAATTGCAAAAAATTTCTTATAACACTGCATTAAATATCATTAATGTACTCTTACCA 60

Qy 61 TATAGTCTTTTGGACAGCTCTGTAGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120

Db 61 AGCATATCTTTTCTGTACTACTATAACAAGATGGTGGGGGTAA-----CTTCTAT 114

Qy 121 ATTAGTCAAGTACAACTCAAGTATATCACACTTTAGAAAATTTCTGTCTGAAGA--AA 178

Db 115 ATTAGTGAAGATGTGTACCAAGTGTCTCACATTTTGGTAGCTTTCAGCTTAAGAGAA 174

Qy 179 CTCCTTAATTAAGAAACAATTTCTCTCACTAAAAAGTTTTCGCACTAAAGAAAGATGGTG 238

Db 175 AGCAAACTCAACTGTTGGAGTTTTCGATTTAAACAATGATTGGGATGGGAAGTCCAACTACT- 233

Qy 239 ATATAACAACAAAAGACGATTTTACAGAGTAGCTCCAGGCAATTTGATTTTCAAATAACT 298

Db 234 -TAAGAATAAACACGCTGACTTTTACTGTTCCAACTATTGTTCCAGATACAGAACATC 292

Qy 299 TAATATCAGGATTTTTCAGGAAGTATTTGTTACTCTATGGACGGACCAAGAAATAGAATTTG 358

Db 293 CATTTCCTAGGTTTTCAGGAGCTATCGTTACTCAATGGGTGGCCCAAGAAATAGAATTCG 352

Qy 359 AAGTCGATATCAACAATTTAATCCAAAACACCGATTAACAATGATGATCTGATTAATGGTG 418

Db 353 AAATATCTTATGAAGCATTCGACGTAAAGTCTTAATATATCAATATCAAAATGACGCGC 412

Qy 419 AATACTATAACATTTTTCGATTTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTAG 478

Db 413 ACAGGTACTGCGCTCTATCTCATCACATCGGAGCCATGGAAGCTGATTAATTTGTCT 472

Qy 479 TACTTAAAAATGACGGCATTAACCTTTTATGTCAATTTGATGGTTTAAATCTTGTCTATGACATTA 538

Db 473 TCTTAAAAACGAAGGGTTAATTGACATATCACTTGCAATAAATGCAATGTTTATGATATA 532

Qy 539 CAGCTGAAGGAGTATCTTTTCGTACATATGATGATGATGATGATGATGATGATGATGATGAT 598

Db 533 TAAATGACAAAGTACCTGTTTCTCTTATATATGCGAGGTATTTGGTACTGATTTGATTT 592

Qy 599 CTATTTTAAAGACCTCAATCTCAAAATTTGCTTACCAAGGAAATAGTATTTAGTTTACC 658

Db 593 CTATGTTTGAAGCTACAGTCTTAAATTTCTTACCAAGGAAATAGTATTTAGTTTACT 652

Qy 659 CTATCACACACAGAGTCTCTGCAATTTATGTTGGTACTTACCATGGCGTTATTTGGTATA 718

Db 653 CTATTAATCCGAAACCTCTGTTTTCATCGTGGGCAATTTCCACAGGATCATAGTAAATG 712

Qy 719 AATTGAGAGATACCTGTTAATACTCTCTGATGATTAATGATGATGATGATGATGATGATGAT 778

Db 713 AGTTTAGAGATATTTCTGCAATAGTACTCAACTCAACTCAACTCAACTCAACTCAACTCAACT 772

Qy 779 CTGCTTCAGTAACTCTTTCGAGTCTTTCGAGGAGAAATTTGGAGTATGAGTTTCACT 838

Db 773 TTGCAACAGTAACACTAAATGTGTCTACCTTTGGTTAGAACTTGGAGGAGATTTAACT 832

Qy 839 TC 840

Db 833 TC 834

RESULT 15

AF082748

LOCUS

DEFINITION Ehrlichia canis strain DJ (North Carolina) 28 kDa outer membrane protein gene, partial cds.

ACCESSION AF082748

VERSION AF082748.1 GI:3769530

KEYWORDS

SOURCE

ORGANISM Ehrlichia canis.

Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Anaplasmataceae; Ehrlichia.

REFERENCE 1 (bases 1 to 834)

McBride, J.W., Yu, X.J. and Walker, D.H.

Molecular cloning of the gene for a conserved major immunoreactive 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic antigen

Clin. Diagn. Lab. Immunol. 6 (3), 392-399 (1999)

JOURNAL

MEDLINE

PUBMED 10225842

REFERENCE 2 (bases 1 to 834)

McBride, J.W., Yu, X.J. and Walker, D.H.

Direct Submission

AUTHORS

TITLE

JOURNAL

FEATURES

source

organism="Ehrlichia canis"



```
/strain="DJ (North Carolina)"
/db_xref="taxon:944"
1. >834
/notes="p28"
/codon_start=1
/transl_table=11
/product="28 kDa outer membrane protein"
/protein_id="AAC64554.1"
/db_xref="GI:3769531"
/translation="MNCKKILITLALISLMYSIPISFSDTIQDNMGNPNYISGKV
PSVSHFGSKBESKSTGVFGLKHDWDGSPILKNKHADFTVNYFRVNNPFLGF
AGAIGYSGGPRIEFISYEAFVKSINYNQNDHRYCALSHSTSAAMEADKVFLLK
NEGLIDISLAINACYDIINDKVPVSPYICAGIGDILSMFEATSPKISYQKGLISYS
INPETSFIGGHPHRIIGNEFRDIPAIVPSNSTITSGPQFATVTLNVCHFGLGLGRF
NF"
BASE COUNT      277 a   147 c   151 g   259 t
ORIGIN
Query Match      28.4%; Score 238.4; DB 1; Length 834;
Best Local Similarity 57.7%; Pred. No. 5.4e-33;
Matches 486; Conservative 0; Mismatches 346; Indels 10; Gaps 3;
Qy 1 ATGAATTATAAGAAAATTCCTAGTAAGAGCGGTTAAATCTCATTAATGTCAATCTTACCA 60
Db 1 ATGAATTGCAAAAATTCCTATACAACTGCATTAATATCAATTAATGTACTCTATTACCA 60
Qy 61 TATCAGCTCTTTTGCAGATCCTGTAGTTCAAGAAGCTAATGATACAAAGAGGCTTCTAC 120
Db 61 AGCATATCTTTTCTGATCTATACAGATGTAACATGGTGGTGTAA-----CTTCTAT 114
Qy 121 ATTAGTCCAAAGTACAAATCCAAAGTATATCACATTTTAGAAAAATCTCTGCTGAAGA--AA 178
Db 115 ATTAGTCCAAAGTATGATACCAAGTCTCACATTTTGTGTAGCTTCTCAGCTAAAGAGAA 174
Qy 179 CTCCTATTAAATGGAACAAATTCCTCACTAAAAAGTTTTCGGACTAAAGAAAGATGGTG 238
Db 175 AGCAATCAACTGTTGGAGTTTGGATTAACATGATGGGATGGAAGTCCCAATACT- 233
Qy 239 ATATAACAAAAAGACGATTTTACAGAGTAGCTCCAGGATTCATTTTCAAAATAACT 298
Db 234 -TAAGAATAAACACGCTGACTTTACTGTTCCAACTATTCTGTCAGATACAGAACATC 292
Qy 299 TAATATCAGGATTTTCAGGAAGTATGTTGTTACTCTATGAGCGGACCAAGATAGAACTTG 358
Db 293 CATTTCTAGGTTTGCAGGAGCTATCGGTTACTCAATGGTGGGCCCAAGATAGAAATCG 352
Qy 359 AAGCTGATATCAACAATTTAATCCNAAAAACCGATAAAGTAAAGTACTGATTAATGGTG 418
Db 353 AAATATCTTAAAGACATTCGACGTAAAGTCCCTAATATCAATTAATCAAAATGACCGC 412
Qy 419 AATACTATAACATTTTGCATTTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTAG 478
Db 413 ACAGTACTGGCTCTATCTCATCACATCGGCAGCCATGGAAGCTGATAAATTTGTCT 472
Qy 479 TACTTAAAAATGACGGCAATACTTTTATGTCAATGATGGTTAATATCTGTCTATGACATTA 538
Db 473 TCTTAAAAAAGCAAGGGTTAATTCACATATCACTTGCAATAAATGCATGTTATGATATA 532
Qy 539 CAGCTGAGGAGTATCTTCGTACCATATGCTGCGAGGTATAGGACAGATCTTATCA 598
Db 533 TAAATGCAAAAGTACCTGTTTCTCCTTATATATGCGCAGGTAATGGTACTGATTTGATTT 592
Qy 599 CTATTTTAAAGACTCAATCTAAATTTGCTTACCAAGGAAAAATAGGTATTAGTTACC 658
Db 593 CTATGTTTGAAGCTACAGTCTTAATTTCTTACCAAGGAAAACTGGCATTTAGTTACT 652
Qy 659 CTATCACACCAAGAGTCTCTGCAATTTATGTTGGTACTACTACCATGGCGTTATTGGTAATA 718
Db 653 CTATTAATCCGGAACCTCTGTTTTCATCGGTGGGCATTTCCACAGGATCATAGGTAATG 712
Qy 719 AATTGAGAGATACCTGTATAACTCTCTGTAGTATTAATGATGCTCCTCAACACCAT 778
Db 713 AGTTTAGAGATATTTCTTGCATATGATCTAGTAACTCAACTACAATAAGTGGACCAACAT 772
```

```
779 CTGCTTCAGTAACCTCTTGACGTTGGATACCTTTGGCGAGAAATTTGGAATGAGGTTCACT 838
773 TTGCAACAGTAACACTAAATGTGTGTCTACCTTTGCTTTAGAACTTTGGAGGAAGATTTAAC 832
839 TC 840
833 TC 834
RESULT 16
AF082749 Ehrlichia canis strain Fuzzy 834 bp DNA linear BCT 26-JUL-2000
LOCUS Ehrlichia canis strain Fuzzy (North Carolina) 28 kDa outer membrane
DEFINITION protein gene, partial cds.
ACCESSION AF082749
VERSION AF082749.1 GI:3769532
KEYWORDS Ehrlichia canis.
SOURCE Ehrlichia canis.
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE 1 (bases 1 to 834)
AUTHORS McBride, J.W., Yu, X.J. and Walker, D.H.
TITLE Molecular cloning of the gene for a conserved major immunoreactive
28-kilodalton protein of Ehrlichia canis: a potential
serodiagnostic antigen
JOURNAL Clin. Diagn. Lab. Immunol. 6 (3), 392-399 (1999)
MEDLINE 99242757
PubMed 10225842
REFERENCE 2 (bases 1 to 834)
AUTHORS McBride, J.W., Yu, X.J. and Walker, D.H.
TITLE Direct Submission
JOURNAL Submitted (07-AUG-1998) Pathology, University of Texas Medical
Branch, 301 University Blvd., Galveston, TX 77555-0609, USA
FEATURES
source
1. 834
/organism="Ehrlichia canis"
/strain="Fuzzy (North Carolina)"
/db_xref="taxon:944"
1. >834
/notes="p28"
/codon_start=1
/transl_table=11
/product="28 kDa outer membrane protein"
/protein_id="AAC64554.1"
/db_xref="GI:3769533"
/translation="MNCKKILITLALISLMYSIPISFSDTIQDNMGNPNYISGKV
PSVSHFGSKBESKSTGVFGLKHDWDGSPILKNKHADFTVNYFRVNNPFLGF
AGAIGYSGGPRIEFISYEAFVKSINYNQNDHRYCALSHSTSAAMEADKVFLLK
NEGLIDISLAINACYDIINDKVPVSPYICAGIGDILSMFEATSPKISYQKGLISYS
INPETSFIGGHPHRIIGNEFRDIPAIVPSNSTITSGPQFATVTLNVCHFGLGLGRF
NF"
BASE COUNT      277 a   147 c   151 g   259 t
ORIGIN
Query Match      28.4%; Score 238.4; DB 1; Length 834;
Best Local Similarity 57.7%; Pred. No. 5.4e-33;
Matches 486; Conservative 0; Mismatches 346; Indels 10; Gaps 3;
Qy 1 ATGAATTATAAGAAAATTCCTAGTAAGAGCGGTTAAATCTCATTAATGTCAATCTTACCA 60
Db 1 ATGAATTGCAAAAATTCCTATACAACTGCATTAATATCAATTAATGTACTCTATTACCA 60
Qy 61 TATCAGCTCTTTTGCAGATCCTGTAGTTCAAGAAGCTAATGATACAAAGAGGCTTCTAC 120
Db 61 AGCATATCTTTTCTGATCTATACAGATGTAACATGGTGGTGTAA-----CTTCTAT 114
Qy 121 ATTAGTCCAAAGTACAAATCCAAAGTATATCACATTTTAGAAAAATCTCTGCTGAAGA--AA 178
Db 115 ATTAGTCCAAAGTATGATACCAAGTCTCACATTTTGTGTAGCTTCTCAGCTAAAGAGAA 174
Qy 179 CTCCTATTAAATGGAACAAATTCCTCACTAAAAAGTTTTCGGACTAAAGAAAGATGGTG 238
Db 175 AGCAATCAACTGTTGGAGTTTGGATTAACATGATGGGATGGAAGTCCCAATACT- 233
Qy 239 ATATAACAAAAAGACGATTTTACAGAGTAGCTCCAGGATTCATTTTCAAAATAACT 298
Db 234 -TAAGAATAAACACGCTGACTTTACTGTTCCAACTATTCTGTCAGATACAGAACATC 292
Qy 299 TAATATCAGGATTTTCAGGAAGTATGTTGTTACTCTATGAGCGGACCAAGATAGAACTTG 358
Db 293 CATTTCTAGGTTTGCAGGAGCTATCGGTTACTCAATGGTGGGCCCAAGATAGAAATCG 352
Qy 359 AAGCTGATATCAACAATTTAATCCNAAAAACCGATAAAGTAAAGTACTGATTAATGGTG 418
Db 353 AAATATCTTAAAGACATTCGACGTAAAGTCCCTAATATCAATTAATCAAAATGACCGC 412
Qy 419 AATACTATAACATTTTGCATTTATCTCGTAAAGATGCAATGGAAGATCAGCAATATGTAG 478
Db 413 ACAGTACTGGCTCTATCTCATCACATCGGCAGCCATGGAAGCTGATAAATTTGTCT 472
Qy 479 TACTTAAAAATGACGGCAATACTTTTATGTCAATGATGGTTAATATCTGTCTATGACATTA 538
Db 473 TCTTAAAAAAGCAAGGGTTAATTCACATATCACTTGCAATAAATGCATGTTATGATATA 532
Qy 539 CAGCTGAGGAGTATCTTCGTACCATATGCTGCGAGGTATAGGACAGATCTTATCA 598
Db 533 TAAATGCAAAAGTACCTGTTTCTCCTTATATATGCGCAGGTAATGGTACTGATTTGATTT 592
Qy 599 CTATTTTAAAGACTCAATCTAAATTTGCTTACCAAGGAAAAATAGGTATTAGTTACC 658
Db 593 CTATGTTTGAAGCTACAGTCTTAATTTCTTACCAAGGAAAACTGGCATTTAGTTACT 652
Qy 659 CTATCACACCAAGAGTCTCTGCAATTTATGTTGGTACTACTACCATGGCGTTATTGGTAATA 718
Db 653 CTATTAATCCGGAACCTCTGTTTTCATCGGTGGGCATTTCCACAGGATCATAGGTAATG 712
Qy 719 AATTGAGAGATACCTGTATAACTCTCTGTAGTATTAATGATGCTCCTCAACACCAT 778
Db 713 AGTTTAGAGATATTTCTTGCATATGATCTAGTAACTCAACTACAATAAGTGGACCAACAT 772
```

```
Db 175 AGCAATCACTGTTGGAGTTTTCGATTAAACATGATTGGATGGAGTCCCAATCT- 233
Qy 239 ATATAACAAAAGAACGATTTTACAGAGTAGCTCCAGGATGATTTCACAAATAACT 298
Db 234 -TAAGAATAAACACGCTGACTTTACTGTTCCAAACTATTTCGTTACAGATACGAGCAATC 292
Qy 299 TAATATCAGGATTTTCAGGAAGTATTGTTACTCTATGGACGGACCAAGATAGAACTTG 358
Db 293 CATTTCTAGGGTTTCAGGAGCTATCGGTTACTCAATGGGTGGCCCAAGATAGAAATTCG 352
Qy 359 AAGCTGCATATCAACAAATTAATCAAAAAACACCGGTAACCAATCATCTGATATGCTG 418
Db 353 AAATATCTTATGAAGCATTGACGCTAAAGATCCTTAATATCAATATCAAAATGACGGC 412
Qy 419 AATACTATAAACATTTTCGATATCTCGTAAGATGCAATGGAAGATCAGCAATATGTAG 478
Db 413 ACAGTACTGCGCTCTATCTCATCACATCGGACCCATGGAAGCTGATAAATTTGTCT 472
Qy 479 TACTTAAAAATGACGGCATAACTTTTATGTCATGATGTTAAATACCTTGTCTATGACATA 538
Db 473 TCTTAAAAAAGCAAGGGTTAATTTGACATATCACTTGCAATAAATGATGTTATGATATA 532
Qy 539 CAGCTGAAGGAGTATCTTTTCGTACCATATGATGTCAGGTATAGGACAGATCTTATCA 598
Db 533 TAATGACAAAGTACTGTTCTCTTATATATGCGCAGGATTTGGTACTGATTTGATTT 592
Qy 599 CTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAATAGGTATTAGTTACC 658
Db 593 CTATGTTTGAAGTACAGTCTCTAAATTTCTCTACCAAGGAAAACTGGGCATTAGTTACT 652
Qy 659 CTATCACACCAAGTCTCTGATTTTATGTTGGATCTACCATGGCGGTTATGTTAATA 718
Db 653 CTATTAATCCGGAAACCTCTGTTTTCATCGGTGGGCATTTCCACAGGATCATAGGTAATG 712
Qy 719 AATTGAGAAGTACTCTGTAATACTCTCTAGTATTAATGATGCTCTCTCAACCAAT 778
Db 713 AGTTTAGAGATATCTCTGATAGTACTAGTAACTCACTACATATAGTGGACCAAT 772
Qy 779 CTGCTTCAGTAACTCTTCAGCTGGATGACTTTGGCGGAAATTTGGAATGAGGTTACCT 838
Db 773 TTGCAACAGTAACTAAATGTTGTCATCTTTGTTTGTAGAACTTGGAGGAGATTTAACT 832
Qy 839 TC 840
Db 833 TC 834
```

```
RESULT 17
AF082750 LOCUS Ehrlichia canis strain Florida 28 kDa outer membrane protein gene,
DEFINITION partial cds.
ACCESSION AF082750
VERSION AF082750.1 GI:3769534
KEYWORDS
SOURCE Ehrlichia canis.
ORGANISM Ehrlichia canis
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE 1 (bases 1 to 834)
AUTHORS McBride,J.W., Yu,X. and Walker,D.H.
TITLE Molecular cloning of the gene for a conserved major immunoreactive
28-kilodalton protein of Ehrlichia canis: a potential
serodiagnostic antigen
JOURNAL Clin. Diagn. Lab. Immunol. 6 (3), 392-399 (1999)
MEDLINE 99242757
PUBMED 10225842
REFERENCE 2 (bases 1 to 834)
AUTHORS McBride,J.W., Yu,X.J. and Walker,D.H.
TITLE Direct Submission
JOURNAL Submitted (07-AUG-1998) Pathology, University of Texas Medical
Branch, 301 University Blvd., Galveston, TX 77555-0609, USA
```

```
FEATURES
source
Location/Qualifiers
1..834
/organism="Ehrlichia canis"
/strain="Florida"
/db_xref="taxon:944"
1..834
/notes="p28"
/codon_start=1
/transl_table=1
/product="28 kDa outer membrane protein"
/protein_id="AAC64556.1"
/db_xref="GI:3769535"
/translating="MNCKILITLISLMYSPISFSDTIOQDNMGNGFYISGVY
PSVSHGSPSAKESKSTGVFGLKHPDWSPLKNKHADFTVPSVGRVNNPLGF
AGAIGSMGPGRIEFSEISAFDVKSNINFDHRYCALSHSISAMEADKVFLLK
NEGILDSLAINACYDILNDKVPVSPYICAGIDTDLISMEFATSPKISYQKGLGISY
INPETSFVIGGHFRIIENFRDIPAIVPSNSTTISGPQFATVTLNCHFLGLELGRF
NF"
BASE COUNT 277 a 147 c 151 g 259 t
ORIGIN
```

```
Query Match 28.4%; Score 238.4; DB 1; Length 834;
Best Local Similarity 57.7%; Pred. No. 5.4e-33;
Matches 486; Conservative 0; Mismatches 346; Indels 10; Gaps 3;
Qy 1 ATGAATTATAAGAAAATCTAGTAAGAAGCGCGTTAATCTCAATTAATGCTCAATCTTACCA 60
Db 1 ATGAATTTGCAAAAAAATTTCTTATAACAACTGCATTAATATATCAATTAATGCTCAATCTTACCA 60
Qy 61 TATCAGTCTTTTTCAGATCCTGTAGGTTCAAGAACTAATGATATAACAAAGAGGCTTCTAC 120
Db 61 AGCATATCTTTTCTGATCTATACAGATGGTAACATGGTGGTAA-----CTTCTAT 114
Qy 121 ATTAGTCAAAAGTACAATCCAAATTCCTCACTAAAAAGTTTCGGACTAAGAAAGATGGTG 238
Db 175 AGCAAAATCAACTGTGGAGTTTTTGGATTAACATGATGGGATGGAGTCCCAATCT- 233
Qy 239 ATATAACAAAAAAGACGATTTTACAAAGATAGCTCCAGGCAATGATTTTCAAAATAACT 298
Db 234 -TAAGAATAACACGCTGACTTTACTGTTCCAACTATTCTGTCAGATACGAGCAATC 292
Qy 299 TAATATCAGGATTTTTCAGGAAGTATTGTTACTCTATGGACGGACCAAGATAGAACTTG 358
Db 293 CATTTCTAGGGTTTCAGGAGCTATCGGTTACTCAATGGGTGGCCCAAGATAGAAATTCG 352
Qy 359 AAGCTGCATATCAACAAATTTAATCCAAAAACCGATTAACAATGATACTGATATGCTG 418
Db 353 AAATATCTTATGAAGCATTCGACGTAAAGTCTCAATATCAATATCAAAATGACGGC 412
Qy 419 AATACTATAACATTTTTCGATTTACTCTGTAAGATGCAATGGAAGATCAGCAATATGTAG 478
Db 413 ACAGTACTGCGCTCTATCTCATCACATCGGACGCCATGGAAGCTGATAAATTTGTCT 472
Qy 479 TACTTAAATGACGGCATAACTTTTATGTCATTTGATGTTTAACTTCTGCTATGACATA 538
Db 473 TCTTAAAAACGAAGGGTTAATTCACATATCACTTGCAATAAATGATGTTTATGATATA 532
Qy 539 CAGCTGAAGGATATCTTCGTACATATGATGTCAGGATAGGATAGGACAGATCTTATCA 598
Db 533 TAAATGACAAAGTACCTGTTTCTCTTATATATGCGCAGGATTTGGTACTGATTTGATTT 592
Qy 599 CTATTTTAAAGACCTCAATCTAAATTTGCTTACCAGGAAAAATAGGTATTAGTTACC 658
Db 593 CTATGTTTGAAGTACAGTCTTAAATTTCTTACCAGGAAAACTGGGCATTAGTTACT 652
Qy 659 CTATCACACCAAGATCTCTGCAATTTATTTGTTGGATGATCTACCATGGCGGTTATGTTAATA 718
Db 653 CTATTAATCCGGAAACCTCTGTTTTCATCGGTGGGCATTTCCACAGGATCATAGGTAATG 712
```

```

Qy 719 AATTTGAGAGTACCTGTAATACTCTGTAGTATTAATGATGCTCCTCAAAACCAT 778
Db 713 AGTTTGTAGATATTCCTGCAATGATACCTAGTAATCACTAACAATAGTGGACCAAT 772
Qy 779 CTGCTTCAGTAACCTCTTCAGCTTGATATCTTTGGCGAGAAATTTGGATGAGGTTCACT 838
Db 773 TTGCAACAGTAACTAAATGTGTCTACTTTGGTTTGAAGCTTTGGAGGAAGATTTAACT 832
Qy 839 TC 840
Db 833 TC 834

RESULT 18
AF393388 1342 bp DNA linear BCT 27-MAR-2002
LOCUS Ehrlichia chaffeensis strain V6 outer membrane protein p28 gene,
DEFINITION complete cds.
ACCESSION AF393388
VERSION AF393388.1 GI:15991529
KEYWORDS
SOURCE
ORGANISM Ehrlichia chaffeensis.
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE
1 (bases 1 to 1342)
Long, S.W., Zhang, X.F., Qi, H., Standaert, S., Walker, D.H. and Yu, X.J.
Antigenic variation of Ehrlichia chaffeensis resulting from
differential expression of the 28-kilodalton protein gene family
Infect. Immun. 70 (4), 1824-1831 (2002)
JOURNAL 21893092
MEDLINE
PUBMED 11895944
REFERENCE
2 (bases 1 to 1342)
Yu, X.-J., Zhang, X.-F. and Walker, D.H.
Direct Submission
AUTHORS
TITLE Submitted
JOURNAL (20-JUN-2001) Pathology, University of Texas Medical
Branch, 301 Univ. Blvd., Galveston, TX 77555-0609, USA
FEATURES
source
1. 1342
/organism="Ehrlichia chaffeensis"
/strain="V1"
/db_xref="taxon:945"
237..1082
/note="p28-19; 28 kDa outer membrane protein"
/codon_start=1
/transl_table=1
/product="outer membrane protein p28"
/db_xref="GI:15991530"
/protein_id="AA12918.1"
/translation="MNYKVFITSLISLPGVSDPAGSGINGNFYISGYMP
SASHFVFSKEERTTVGFLKQNDGSAISNSPNDVTVSNYSFKYENNPFLGF
AGAIGYMDGPRILEVSYETFDVKQNNYKNEAHRYCALSHNSAADMSSANNVF
LKNEGLDISPMLNACVDVVEGIPSPYICAGIGTDLVSMFEATNPKISYQKLGIS
YSISPEASVFIGHFHKNVINEFRDITPIITGSTLAGKNYPAIVLDVCHFGIEUG
GRFAP"
BASE COUNT 437 a 215 c 242 g 448 t
ORIGIN
Query Match 27.9%; Score 234; DB 1; Length 1342;
Best Local Similarity 58.2%; Pred. No. 2.8e-32;
Matches 496; Conservative 0; Mismatches 335; Indels 21; Gaps 4;
Qy 1 ATGAATTATAGAAAATTTAGTAGAAGCGCGTTAATCTCATTAATGTCAATCTTACCA 60
Db 237 ATGAATTACAAAAGTTTTCATAACAAGTCGATGATATCAATTAATATCTTCTACCT 296
Qy 61 TATCAGTCTTTTGGAGATCTGTAGTTCAGAACTAATGATAACAAAGAGGCTTCTAC 120
Db 297 GGAGTATCATTTCCGACCCAGCAGGTAGTGGTATTAAACGGTAAT-----TTCTAC 347
Qy 121 ATTAGTCGAAGTACATCCCAAGTATATCACTTTAGAAAATTTCTGCT---GAAGAA 177
Db 348 ATCAGTGAAAATACATGCCAAGTGCTTCGCAATTTTGGAGTATTTCTCTGCTAAGGAAGAA 407

```

```

Qy 178 ACTCTATTAAATGGAAACAAATTTCTCTCACTAAAAAAGTTTTCGGACTAAGAAAGATGGT 237
Db 408 AGAAATCAACAGTTGGAGTGTGGAGCTGAAGCAAAATTTGGGACGGAAGCGCAATATCC 467
Qy 238 GATATAACAAAAAAGACGATTTTACAGAGTAGCTCCAGGCATTGATTTTCAAAAATAC 297
Db 468 AACTCCCTCCCAACAGATGATTCACACTGTCTCAAAATTTATTCATTTAATATGAACAAC 527
Qy 298 TTAATATCAGATTTTTCAGGAAGTATTGGTTACTCTATGACGAGCAACCAAGATAGAACTT 357
Db 528 CCGTTTTAGGTTTTCAGGAGCTATTGGTTACTCAATGATGCTCCAGAAATAGAGCTT 587
Qy 358 GAAGCTGCATATCAACAATTTAATCCAAAAACCAACCGATAACATGATCTGATTAATGGT 417
Db 588 GAAGTATCTTATGAACATTTGATGTAAAAAATCAAGGTAAACAAATTTAAGAAAGCA 647
Qy 418 GAATACATATAACATTTTGGCAATATCTCGTAAGATGCAATG-----GAAGATCAGCAA 471
Db 648 CATAGATATTGTGCTCTATCCCAATCACTCAGCAGCAGACATAGTAGTGCAGTAATAAT 707
Qy 472 TATGTAGTACTTAAAAATGACGCATAAATTTTATGTCTATTGATGGTTAATACTTGTCTAT 531
Db 708 TTGTCTTTCTAAAAATGAAGATTAATTGACATATCATTTATGCTGAACCGATGCTAT 767
Qy 532 GACATTAACAGCTGAAGGAGTATCTTTGCTACCATATGATGTCAGGTATAGGAGCAGAT 591
Db 768 GACGTAGTAGCGAAGGCATACCTTTTCTCTTATATATGCGCAGGTATCGGTACTGAT 827
Qy 592 CTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAGAGAAAATAGGTATT 651
Db 828 TTAGTATCCATGTTTGAAGCTACAAATCTTAAAAATTTCTTACCAAGGAAAGTTAGGTTA 887
Qy 652 AGTTTACCCTATCACACAGAAAGTCTCTGCATTTTATGCTGTGATACCTACCATGGCGTTATT 711
Db 888 AGCTACTCTATAGCCCGAAGCTTCTGTGTTTATTGTTGGGCACTTTTCATAAGGTAAATA 947
Qy 712 GGTAATAAATTTGAGAAGATAC---CTGTAATAACTCCTCTGTAGTATTAAATGATGCTCCT 768
Db 948 GGGAAACAAATTTAGAGATATTCCTACTATTAATACCTACTGGATCAACACTTGCAGGAAA 1007
Qy 769 CAACACCATCTGCTTCAGTAACTCTTGACGTGGATGATCTTTGGCGGAGAAAATGGAGATG 828
Db 1008 GGAACCTACCTCGCAATAGTAATACCTGATGTATGCCACTTTTGGAAATAGAACTTGGAGGA 1067
Qy 829 AGGTTCACTTC 840
Db 1068 AGGTTTCTTTC 1079

```

```

RESULT 19
AF393392 1315 bp DNA linear BCT 27-MAR-2002
LOCUS Ehrlichia chaffeensis strain V6 outer membrane protein p28 gene,
DEFINITION complete cds.
ACCESSION AF393392
VERSION AF393392.1 GI:15991537
KEYWORDS
SOURCE Ehrlichia chaffeensis.
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE
1 (bases 1 to 1315)
Long, S.W., Zhang, X.F., Qi, H., Standaert, S., Walker, D.H. and Yu, X.J.
Antigenic variation of Ehrlichia chaffeensis resulting from
differential expression of the 28-kilodalton protein gene family
Infect. Immun. 70 (4), 1824-1831 (2002)
JOURNAL 21893092
MEDLINE
PUBMED 11895944
REFERENCE
2 (bases 1 to 1315)
Yu, X.-J. and Zhang, X.-F.
Direct Submission
AUTHORS
TITLE Submitted (20-JUN-2001) Pathology, University of Texas Medical
JOURNAL

```

Branch, 301 Univ. Blvd., Galveston, TX 77555-0609, USA

FEATURES  
source

Location/Qualifiers

1..1315  
/organism="Ehrlichia chaffeensis"  
/strain="V6"  
/db\_xref="taxon:945"  
238..1083  
/note="p28-19; 28 kDa outer membrane protein"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane protein p28"  
/protein\_id="AAL12922.1"  
/db\_xref="GI:15991538"  
/translation="MNYKKVITSALISLSLPGVSFSDPAGSINGNFVISKYMP  
SASHGVPSAKEERTVTVGLKQNWDSALNSNSPNDVFTVNSYKYNPFLGF  
AGATGYSDMPRIELEVYFEDVKNQNNYKSEAHRYCALSHNSADMSANNFV  
LKNELLDISPMILNACVGVGEPFSPYICAGICTDLVSFEATNPRIKISYQGLGLS  
YSISPEASVFIGGHPHKVIGNEFRDPTIIPGTSLAEKGNVPAIVLDVCHFGIELG  
GRFAP"

CDS

BASE COUNT 428 a 205 c 238 g 444 t

ORIGIN

Query Match .27.8%; Score 233.2; DB 1; Length 1315;  
Best Local Similarity 59.1%; Pred. No. 3.9e-32;  
Matches 507; Conservative 0; Mismatches 318; Indels 33; Gaps 5;

QY 1 ATGAATTATAGAAAATCTAGTAGAGAGCGGTTAATCTCTAATGTCATCTTACCA 60  
DB |||||  
QY 238 ATGAATACAAAAAGTTTCATACAAAGTGCATTGATCATTAATATCTCTCTACCT 297  
DB |||||  
QY 61 TATCAGTCTTTTGCAGATCCTGTAGTTCACAGAACTAATGATTAACAAAGAGGCTTCTAC 120  
DB |||||  
QY 298 GGAGTATCATTTCCGACCCAGCAGTAGTGTATTAACGGTAAT-----TTCCTAC 348  
DB |||||  
QY 121 ATTAGTCAAAAGTACAAATCTAGTAGAGAGCGGTTAATCTCTGCT---GAAGAA 177  
DB |||||  
QY 349 ATCAGTGGAAAATACATGCCAAGTCTTCGATTTTGGAGTATCTCTGCTAAGGAAGA 408  
DB |||||  
QY 178 ACTCCTATTATGGAACAAATCTCTACTAAAAAGTTTTCGGACTAAAGAAAGATGGT 237  
DB |||||  
QY 409 AGAAATACACAGTTTGGAGTGTTCGACTGAAGCAAAATTTGGACGGAAGCGCAATATCC 468  
DB |||||  
QY 238 GATATAACAAAAAGAGATTTTACAGAGTAGTCCAGGCGATGATTTTCAAAATAAC 297  
DB |||||  
QY 469 AACTCCTCCCAACAGATGATCTACTGTCTCAAAATATTCATTAATATGAAAAACAC 528  
DB |||||  
QY 298 TTAATATCAGGATTTTCAGGAGTATTTGGTACTCTATGGACGGACCAAGAAATAGAACTT 357  
DB |||||  
QY 529 CCGTTTTTAGGTTTTGCAGGAGCTATTGTTACTCAATGGATGGTCCAAGAAATAGAGCTT 588  
DB |||||  
QY 358 GAAGCTGCATATCACAATTTAATCCAAAAACACCGATAACATGATGATGATGATGAT 417  
DB |||||  
QY 589 GAAGTATCTTATGAACATATTTGATGTAAAAAATCAAGGTAAACA-----TTATAAGAGT 642  
DB |||||  
QY 418 GAATACTATAACATTTTGCATTATCTCGTAAGATGCA-----ATGGAAGAT 465  
DB |||||  
QY 643 GAAGCACATAGATATTGTGCTCTATCCCAATCTCAGCAGCAGACATGATGATGCAAGT 702  
DB |||||  
QY 466 CAGCAATATGTAGTACTTTAAAAATGACGCGCATAACTTTTATGTCATTGATGTTAATACT 525  
DB |||||  
QY 703 AATAATTTGCTTTCTAAAAAATGAAGGATTACTTGACATATCATTTATGCTGAACGCA 762  
DB |||||  
QY 526 TGCTATGACATTAACAGTGAAGGATATCTTCGTACCATATGATGATGTCAGGATATAGA 585  
DB |||||  
QY 763 TGCTATGACGTATGAGCGAAGGCATACCTTTTCTCTTATATATGCGCAGGATTCGGT 822  
DB |||||  
QY 586 GCAGATCTTATCAGTATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGAAAAATA 645  
DB |||||  
QY 823 ACTGATTTAGTATCCATGTTTGAGAGCTACAAATCTCAAAATTTCTTACCAAGGAAAGTTA 882  
DB |||||  
QY 646 GGTATTAGTATCCCTATCACACAGAGGTCCTCGCATTTTATGGTGGATCTACCATGGC 705  
DB |||||  
QY 883 GGTTTAAGCTACTCTATAAGCCAGAGGCTTCTGTGTTTATTTGGTGGGCACTTTTCATAAG 942  
DB |||||

QY 706 GTTATTGGTAATAAATTTGAGAAGATAC---CTGTAATAACTCTCTAGTAGTAAATGAT 762  
DB |||||  
QY 943 GTAATAGGGAACGAATTTAGAGATATTTCTACTATAATACCTACTGGATCAACTTCCA 1002  
DB |||||  
QY 763 GCTCCTCAAAACCATCTGCTTCACTAATCTTTCAGCTTGGATATCTTTGGCGGAGAAAT 822  
DB |||||  
QY 1003 GAAAAAGGAACACTACCTGCAATAGTAATCTGATGTATGCCACTTTGGATAGAACTT 1062  
DB |||||  
QY 823 GGAATGAGGTTTCACTTC 840  
DB |||||  
QY 1062 GGAGGAAGGTTTGTCTTC 1080  
DB |||||

RESULT 20

AX042315

LOCUS

DEFINITION

AX042315

ACCESSION

AX042315

VERSION

AX042315.1

KEYWORDS

SOURCE

ORGANISM

Ehrlichia chaffeensis.

Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

Anaplasmataceae; Ehrlichia.

1 (bases 1 to 830)

Barbet,A.F., Bowie,M.V., Ganta,R.R., Burrige,M.J., Mahan,S.M.,

McGuire,T.C., Rurangirwa,F.R., Moreland,A.B., Simbi,B.H.,

Whitmire,W.W. and Alleman,A.R.

Nucleic acid vaccines against rickettsial diseases and methods of

use

Patent: WO 0065063-A 11 02-NOV-2000;

UNIVERSITY OF FLORIDA (US)

Location/Qualifiers

1..830

/organism="Ehrlichia chaffeensis"

/db\_xref="taxon:945"

BASE COUNT 278 a 141 c 160 g 251 t

ORIGIN

Query Match 27.2%; Score 228.6; DB 6; Length 830;

Best Local Similarity 58.2%; Pred. No. 3e-31;

Matches 487; Conservative 0; Mismatches 329; Indels 21; Gaps 4;

QY 1 ATGAATTATAGAAAATCTAGTAGAGAGCGGTTAATCTCTAATGTCATCTTACCA 60

DB |||||

QY 3 ATGAATTAACAAAAAGTTTTCATACCAAGTGCATTGATCATTAATATCTCTCTACCT 62

DB |||||

QY 61 TATCAGTCTTTTGCAGATCCTGTAGTTCACAGAACTAATGATTAACAAAGAGGCTTCTAC 120

DB |||||

QY 63 GGAGTATCATTTTCGACCCAGCAGGTAGTGTATTACGGTAAT-----TTCCTAC 113

DB |||||

QY 121 ATTAGTCAAAAGTACAATCTCAAGTATATCAACATTTTAGAAAAATTTCTCTGCT---GAAGAA 177

DB |||||

QY 114 ATCAGTGAAAATACATGTCACCAAGTCTTCGATTTTGGAGTATTTCTCTGCTAAGGAAGA 173

DB |||||

QY 178 ACTCCTATTATAGAAACAAATCTCTCACTAAAAAGTTTTCGGACTTAAGAAAGATGGT 237

DB |||||

QY 174 AGAAATACACAGTTTGGAGTGTTCGACTGAAGCAAAATTTGGGACGGAAGCGCAATATCC 233

DB |||||

QY 238 GATATAACAAAAAGAGAGATTTTACAGAGTAGTCTCCAGGCAATGATTTTCAAAATAAC 297

DB |||||

QY 234 AACTCTCCCAACAGATGATTCCTGTCTCAATATTTCATTTAATATGAAAAACAC 293

DB |||||

QY 298 TTAATATCAGGATTTTTCAGGAAGTATTGGTTACTCTATGGACGGACCAAGAAATAGAACTT 357

DB |||||

QY 294 CCGTTTTTAGGTTTTTCAGGAGGCTATTGGTTACTCAATGGATGGTCCCAAGAAATAGAGCTT 353

DB |||||

QY 358 GAAGCTGCATATCACAATTTAATCCAAAAACACCGATCAATGATGATGATGATGATGAT 417

DB |||||

QY 354 GAAGTATCTTATGAACATTTTATGATGAAAAATCAAGGTAAACAAATTAAGAAATGAAGCA 413

DB |||||

QY 418 GAATACTATAAACATTTTTCGATTATCTCGTAAAGATGCAATG-----GAAGATCAGCAA 471

DB |||||

```

Db      414 CATAGATATGTGCTCTATCCATNACTCAGCAGCAGCATGAGTAGTGCAGTAATAAT 473
Qy      472 TATGTAGTACTTAAAAATGAGGCATAACTTTTATGTCAATGATGGTTAAATACTTGCTAT 531
Db      474 TTTGCTCTTCAAAAAATGAAGGATTACTTGACATATCATTTATGCTGAACGCATGCTAT 533
Qy      532 GACATTACAGCTGAAGGAGTATCTTTCTGTACATATGATGTGAGGATATAGGAGCAGAT 591
Db      534 GACGTAGTAGGCGAAGGCATACCTTTTCTCTATATATGCGCAGGTATCGGTACTGAT 593
Qy      592 CTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAATAGGTATT 651
Db      594 TTAGTATCCATGTTTGAAGCTACAAATCTTAAATTTCTTACCAAGGAAATAGGTATT 653
Qy      652 AGTTACCCATCACACAGAACTCTCTGCAATTTATTTGGTGATACTACCATGGCGTTATT 711
Db      654 AGCTACTCTATAAGCCAGAGCTCTGCTGTTTATTTGGTGCGCACTTTCATAAGGTAATA 713
Qy      712 GGTATAAATTTGAGAGATAC---CTGTATAACTCTCTGTAGTATTAAATGATGCTCCT 768
Db      714 GGGAAACGAATTAGAGATATTTCTACTATAATACCTACTGATCAACACTTGCAGGAAA 773
Qy      769 CAACACCATCTGCTTCAGTAACTCTTGACCTTGATCTTTGGCGGAGAAATTGGA 825
Db      774 GGAAACTACCTGCAATAGTAATACTGATGTATGCCACTTTTGAATAGAAATGGA 830

RESULT 21
LOCUS   AF062761
DEFINITION Ehrlichia chaffeensis 28 kDa major surface antigen multi-gene
         locus, partial sequence.
ACCESSION AF062761
VERSION   AF062761.1 GI:3327958
KEYWORDS
SOURCE   Ehrlichia chaffeensis.
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
         Anaplasmataceae; Ehrlichia.
REFERENCE 1 (bases 1 to 4683)
AUTHORS  Reddy G.R., Sulseon, C.R., Barbet, A.F., Mahan, S.M., Burridge, M.J.
         and Alleman, A.R.
TITLE    Molecular characterization of a 28 kDa surface antigen gene family
         of the tribe Ehrlichiae
JOURNAL  Biochem. Biophys. Res. Commun. 247 (3), 636-643 (1998)
MEDLINE  98321180
PUBMED   9647746
REFERENCE 2 (bases 1 to 4683)
AUTHORS  Reddy, G.R., Sulseon, C.R., Barbet, A.F., Mahan, S.M., Burridge, M.J.
         and Alleman, A.R.
TITLE    Direct Submission
JOURNAL  Submitted (02-MAY-1998) Diagnostic Medicine/Pathobiology, College
         of Veterinary Medicine, Kansas State University, 1800 Denison Ave,
         Manhattan, KS 66506, USA
FEATURES             Location/Qualifiers
     source            1..4683
                     /organism="Ehrlichia chaffeensis"
                     /strain="Arkansas"
                     /db_xref="taxon:945"
     CDS               1..134
                     /note="similar to Cowdria ruminantium MAP1 gene; ORF 1;
                     silent copy"
                     /codon_start=3
                     /evidence=not_experimental
                     /transl_table=11
                     /product="28 kDa major surface antigen-1"
                     /protein_id="AAC26717.1"
                     /db_xref="GI:3327960"
                     /translation="NEFRDISTLKAFAFPSSAATPDLATVTLVCHFGVELGGGFNF"
     terminator        179..188
                     /note="palindromic sequence part 1"
                     /evidence=not_experimental

```

```

terminator           203..212
                     /note="palindromic sequence part 2"
                     /evidence=not_experimental
misc_feature          226..239
                     /note="g-rich sequence"
-35_signal            349..354
                     /evidence=not_experimental
-10_signal            375..380
                     /evidence=not_experimental
RBS                  430..434
                     /evidence=not_experimental
CDS                  439..1299
                     /note="similar to Cowdria ruminantium MAP1 gene; ORF 2;
                     silent copy"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="28 kDa major surface antigen-2"
                     /protein_id="AAC26718.1"
                     /db_xref="GI:3327961"
                     /translations="MNCKEFTTALTLLMSFLPGLSDPVDQDNISGNFYISGKYM
                     PSASHFGVFSAKEERTTVGVFGTIEDWDRCVISRTLTSDIFTVPNYSFKYENNLFSG
                     FAGAIGYMDGPRIELEVSVEAFDVKQGNKYKNEAHRYVALSHLLCTEQIDGAGSA
                     SVLLIENGLDKSFMNLNACVDVISEGIPFSPYCAGIGIDLVNFEAINPKISQGLK
                     GLSYPSPEASVFIGGHFKVIGNEFRDITPIMPESALAGKGNYPATVLDVYFGL
                     ELGGRFNQL"
                     /evidence=not_experimental
                     /note="palindromic sequence part 1"
                     /evidence=not_experimental
terminator           1344..1353
                     /note="palindromic sequence part 2"
                     /evidence=not_experimental
misc_feature          1347..1361
                     /note="g-rich sequence"
-35_signal            1473..1479
                     /evidence=not_experimental
-10_signal            1498..1504
                     /evidence=not_experimental
RBS                  1554..1558
                     /evidence=not_experimental
CDS                  1563..2399
                     /note="similar to Cowdria ruminantium MAP1 gene; ORF 3;
                     silent copy"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="28 kDa major surface antigen-3"
                     /protein_id="AAC26719.1"
                     /db_xref="GI:3327962"
                     /translations="MNCKEFTTALTALVLSMFLPGLSDFSDVQDNISGNFYISGKYM
                     PSASHFGVFSAKEERTTVGVFGTIEDWDRCVISRTLTSDIFTVPNYSFKYENNLFSG
                     AGAIGYMDGPRIELEVSVEAFDVKQGNKYKNEAHRYVALSHLLCTEQIDGAGSA
                     KSGLLDISFMLNACVDVISEGIPFSPYCAGIGIDLVNFEAINPKISQGLKGLSY
                     SINPEASVFIGGHFKVIGNEFRDITPMPESALAGKGNYPATVLDVYFGL
                     NF"
                     /evidence=not_experimental
                     /note="palindromic sequence part 1"
                     /evidence=not_experimental
terminator           2444..2456
                     /note="palindromic sequence part 2"
                     /evidence=not_experimental
misc_feature          2491..2505
                     /note="g-rich sequence"
-35_signal            2618..2624
                     /evidence=not_experimental
-10_signal            2644..2649
                     /evidence=not_experimental
RBS                  2699..2703
                     /evidence=not_experimental
CDS                  2708..3550
                     /note="similar to Cowdria ruminantium MAP1 gene; ORF 4;
                     silent copy"
                     /codon_start=1

```

```

/evidence=not_experimental
/transl_table=11
/product="28 kDa major surface antigen-4"
/protein_id="AAC26720.1"
/db_xref="GI:3327963"
/translation="MNCKKFFITTTLSLMSFLPGISFSDAVQNDVNGNFIYISGKVV
PSVHFGVFSKQERNTTIGVFLKQDWDGSTISKNSPENTFNPVNSFKYNNPFLG
PAGVGLVLMNGPRIELENSETYFDVKQGNKYKDAHYALYTHNSGGKLSNAGDKFV
FLKNEGLDLSIMLNACVDVISEGIPFSPYICAGVGTDLISMFENAIKPKISYQKGLG
SYSISPEASVFGVGHFKVIGNEFRDIPAMIPSTLTGNHFTITVLSVCHFGVELGG
RFNF"
terminator
3586..3596
/feature="palindromic sequence part 1"
/evidence=not_experimental
3610..3622
/feature="palindromic sequence part 2"
/evidence=not_experimental
3643..3655
/feature="g-rich sequence"
3767..3772
/evidence=not_experimental
3788..3793
/evidence=not_experimental
3847..3851
/evidence=not_experimental
3856..>4683
/feature="similar to Cowdria ruminantium MAP1 gene; ORF 5;
expressed copy"
/codon_start=1
/transl_table=11
/product="28 kDa major surface antigen-5"
/protein_id="AAC26716.1"
/db_xref="GI:3327959"
/translation="MNYKKVFTTSALISLISLPGVSPDPAAGSNGNFIYISGKYP
SASHFVFSKQERNTTIGVFLKQDWDGSTISKNSPENTFNPVNSFKYNNPFLG
AGVGLVLMNGPRIELENSETYFDVKQGNKYKDAHYALYTHNSGGKLSNAGDKFV
FLKNEGLDLSIMLNACVDVISEGIPFSPYICAGVGTDLISMFENAIKPKISYQKGLG
SYSISPEASVFGVGHFKVIGNEFRDIPAMIPSTLTGNHFTITVLSVCHFGVELGG
RFNF"
BASE COUNT 1578 a 739 c 833 g 1533 t
ORIGIN
Query Match 27.2%; Score 228.6; DB 1; Length 4683;
Best Local Similarity 58.2%; Pred. No. 1.7e-31;
Matches 487; Conservative 0; Mismatches 329; Indels 21; Gaps 4;
QY 1 ATGAATTATAAGAAATCTAGTAAAGACGGGTAAATCTCATTAATGCTCAATCTTACCA 60
DB 3856 ATGAATTACAAAAGTTTTCATAACAAGTGCATTGATATCATTAAATATCTTCTACCT 3915
QY 61 TATCAGTCTTTTGCAGATCTGTAGGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120
DB 3916 GGAGTATCATTTTCCGACCCAGCAGGTAGTGTATTAACGGTAAT-----TTCTAC 3966
QY 121 ATTAGTCAAAAGTACAATCCAAGTATATCACACTTTAGAAAATCTCTGCT---GAAGAA 177
DB 3967 ATCAGTGAATAATCATCCCAAGTCTTCGCATTTTGGAGTATCTCTGTAAGGAAGA 4026
QY 178 ACTCCTATTATGAACAAATCTCTCACTAAAAAAGTTTTCGGACTAAAGAAAGATGGT 237
DB 4027 AGAATAACAAGTTGGAGTGTGGTGGACTGAAGCAAAATTTGGACGGAGCGCAATATCC 4086
QY 238 GATATAACAAGAGAGATTTTACAGAGTAGTCTCCAGGATTTGATTTTCAAAATAC 297
DB 4087 AACTCTCTCCCAACAGATGATTTCACTGTCTCAAAATTTATTCATTTAAATATGAACAAAC 4146
QY 298 TTAATATCAGATTTTTCAGGAGTATTTGTTACTCTATGAGCGGACCAAGAAATGAACTT 357
DB 4147 CCGTTTTAGGTTTTCAGGAGCTATTTGTTACTCAATGATGTTCCCAAGATAGGTT 4206
QY 358 GAAGCTGCATATCAAAATTTTAAATCCAAAAACACCGGATAACAATGATCTGATAATGGT 417
DB 4207 GAAGTATCTTATGAACAACTTTGATGTATAAAAAATCAAGGTAACTAATTAAGAATGAAGCA 4266

```

```

QY 418 GAATACATAAACATTTTGCATTATCTCGTAAGATCAATG-----GAAGATCAGCAA 471
DB 4267 CATAGATATTGTGCTCTATCCCATAACTCAGCAGCAGACATAGTAGTGCAGTAATAAT 4326
QY 472 TATGTAGTACTTAAAAATGACGGCATAAATTTTATGTCATTGATGGTTAATATCTTGTAT 531
DB 4327 TTTGTCTTTCTAAAAATGAAGNATTAATCTGACATATCATTTATGCTGAACGATGCTAT 4386
QY 532 GACATTACAGCTGAAGAGATATCTTTTCGTAACCATATCATGTGCGAGTATAGGAGCAGAT 591
DB 4387 GACGTAGTAGCGAAGCATACCTTTTCTCTTATATATGCGCAGGTATCGGTACTGAT 4446
QY 592 CTATACATATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGCAAAATAGGTATT 651
DB 4447 TTAGTATCCATGTTTGAAGCTACAAATCTTAAATTTCTTACCAAGGAAATAGGTATT 4506
QY 652 AGTTACCTATCACACCAGAGTCTCTGCATTTTATGTTGATACCTACCATGGCGTTATT 711
DB 4507 AGCTACTCTATAGCCCGAGAGCTTCTGTGTTTATGTTGGCACTTTCATAAGGTAATA 4566
QY 712 GGTAAATAAATTTGAGAAGATAC---CTGTAATAACTCTCTGTAGTATTAATGATGCTCCT 768
DB 4567 GGGAAACGAATTTAGAGATATTTCTTACTATAATACTACTGATCAACACTTGCAGGAAA 4626
QY 769 CAACACCATCTGCTCAGTAACTCTTGCAGTTGGATACCTTGGCGGAGAAATTTGA 825
DB 4627 GGAAACTACCTGCAATAGTAATACTGATGTATGCCACTTTGGAATAGAAATCGGA 4683

RESULT 22
AF393389 LOCUS 1280 bp DNA linear BCT 27-MAR-2002
DEFINITION Ehrlichia chaffeensis strain V2 outer membrane protein p28 gene,
complete cds.
ACCESSION AF393389
VERSION AF393389.1 GI:15991531
KEYWORDS
SOURCE Ehrlichia chaffeensis.
ORGANISM Ehrlichia chaffeensis
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE 1 (bases 1 to 1280)
AUTHORS Long, S.W., Zhang, X.F., Qi, H., Standeert, S., Walker, D.H. and Yu, X.J.
TITLE Antigenic variation of Ehrlichia chaffeensis resulting from
different expression of the 28-kilodalton protein gene family
JOURNAL Infect. Immun. 70 (4), 1824-1831 (2002)
MEDLINE 21893092
PUBMED 11895944
REFERENCE 2 (bases 1 to 1280)
AUTHORS Yu, X.-J., Zhang, X.-F. and Walker, D.H.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2001) Pathology, University of Texas Medical
Branch, 301 Univ. Blvd., Galveston, TX 77555-0609, USA
FEATURES
source
1..1280
/organism="Ehrlichia chaffeensis"
/strain="V2"
/db_xref="taxon:945"
242..1069
/feature="p28-19; 28 kDa outer membrane protein"
/codon_start=1
/transl_table=11
/product="outer membrane protein p28"
/protein_id="AAL12919.1"
/db_xref="GI:15991532"
/translation="MNYKKVFTTSALISLISLPGVSPDPAAGSNGNFIYISGKYP
SASHFVFSKQERNTTIGVFLKQDWDGSTISKNSPENTFNPVNSFKYNNPFLG
PAGVGLVLMNGPRIELENSETYFDVKQGNKYKDAHYALYTHNSGGKLSNAGDKFV
FLKNEGLDLSIMLNACVDVISEGIPFSPYICAGVGTDLISMFENAIKPKISYQKGLG
SYSISPEASVFGVGHFKVIGNEFRDIPAMIPSTLTGNHFTITVLSVCHFGVELGG
RFNF"
BASE COUNT 428 a 207 c 219 g 426 t
ORIGIN

```

```

Query Match      27.2%; Score 228.2; DB 1; Length 1280;
Best Local Similarity 58.0%; Pred. No. 3.1e-31;
Matches 489; Conservative 0; Mismatches 333; Indels 21; Gaps 4;

Qy 1 ATGATTATAAGAAATCTAGTAGAAGCGGTTAAATCTCATTAATCTCAATCTTACCA 60
Db 242 ATGATTACAAAAGTTTTCATACAAGTGCAATGATCATCTATATCTTCTTACCT 301

Qy 61 TATCAGTCTTTTGCAGATCTGTAGGTTCAAGAACTAATGATAACAAAGAGGCTTTCAC 120
Db 302 GGAGTATCATTTTCGACCCCAACAGGTAGTGTATTAACGGCAAT-----TTCTAT 352

Qy 121 ATTAGTCAAGATACAAATCCAAAGTATATACATCTTTAGAAAATTTCTCTGCT---GAAGAA 177
Db 353 ATCAGTGAAAATATATGCCAAGCGCTTCGCAATTTTGGGGTGTCTTCTGCTTAAGGAAGAA 412

Qy 178 ACTCCTATTATATGGAACAAATCTCTCACTAAAGAAATTTTCGGACTAAAGAAAGATGGT 237
Db 413 AGAAATACACAGCTGGAGTATTTGGATGAGCAAGATTTGGATGGCAGTGCAATACCT 472

Qy 238 GATATAACAAAAGAACAGATTTTACAAGAGTAGCTCCAGGCAATTTGATTTTCAAAAATAC 297
Db 473 CACACTCAGAGTAGTACTCCATTTACTCTTTCAACTACTCTATTAATATGAAATAT 532

Qy 298 TTAATATCAGGATTTTTCAGGAAGTATTTGGTTACTCTATGAGCGGACCAAGAAATAGAACTT 357
Db 533 CCATTTCTAGGTTTTCAGGAGCTATTTGGTACTCAATGGATGGTCCCAAGAAATAGAGCTT 592

Qy 358 GAAGCTGCATATCAACATTTAAATCCAAAACACCGATACATGATGATCTGATTAATGGT 417
Db 593 GAAGTATCTTATGAGACATTTGATGTTTAAATCAAGTAAACAA-----CTATAAGAAC 646

Qy 418 GAATACTATAAACATTTTGCATATCTCCTAAAGATGCAATGGAAGATCAGCAATATGTA 477
Db 647 GAAGCTCATAGATCTGTGCTTTATCCGTCAGGCCACAAATGACAGATAATTAATTTGTT 706

Qy 478 GTACTTAAATATGAGCGCATAACTTTTATGTCATGATGGTTAAATCTTGTCTATGACAT 537
Db 707 TTTCTTAAATAATGAAGGCTACTTGACATATCATTTATGCTGAACGCAATGCTATGATGA 766

Qy 538 ACAGCTGAAGGATCTTTTCTGATCATGATGCGAGTATGAGGATGAGTCTTATC 597
Db 767 ATAAGTGAAGAAATACCTTTTCTCCTTACATATATGTCAGGATTTGGTACTGATTA 826

Qy 598 ACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGAAAATAGGTATTAGTTAC 657
Db 827 TCCATGTTGAAGTTACAAACCTTAAATTTCTTACCAAGAAAGTTAGTTAGTTAC 886

Qy 658 CCTATCACACAGAGTCTCTGCAATTTATTTGGTGGATATCACTACCAAGGCGTTATTGGTAAT 717
Db 887 TCTATAAGCCCAAGAACTCTGCTTTGTCGGCGGACATTTCCATAAGGTAATGGGAAC 946

Qy 718 AAATTTGAGAGATACCTGTAATTAATCTCTGATGATTAATGATGCTCTCAACACCA 777
Db 947 GAAATCAGAGATTTCTGCTGATATAATACCTAGTGGATCAAAATCTTACAGGGAATCAC--- 1003

Qy 778 TCTGCTTCAGTAATCTTTCAGCTTGGATATCTTTGGCGGAGAAATTTGGAATCAGGTTTACC 837
Db 1004 CCGCAATAGTAACACTAGATGATGCCACCTTTGGATAGAACTTTGGAGGAAGATTTGCT 1063

Qy 838 TTC 840
Db 1064 TTC 1066

```

```

RESULT 23
AF355200
LOCUS
DEFINITION Cowdria ruminantium Ball-3 major antigenic protein 1 gene, complete cds.
ACCESSION AF355200
VERSION AF355200.1 GI:13448814

```

```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
CDS
BASE COUNT
ORIGIN

```

```

Ehrlichia ruminantium.
Ehrlichia ruminantium
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
1 (bases 1 to 863)
Allsopp,M.T., Dorfling,C.M., Maillard,J.C., Bensaid,A.,
Haydon,D.T., van Heerden,H. and Allsopp,B.A.
Ehrlichia ruminantium major antigenic protein gene (map1) variants
are not geographically constrained and show no evidence of having
evolved under positive selection pressure
J. Clin. Microbiol. 39 (11), 4200-4203 (2001)
21539003
11682561
2 (bases 1 to 863)
Bensaid,A., Allsopp,M.T., Maillard,J.-C., Chantal,I. and Allsopp,B.
Direct Submission
Submitted (02-MAR-2001) Elevage et Medecine Veterinaire, Centre de
Cooperation Internationale en Recherche Agronomique pour le
Developpement, Domaine Duclos, Petit Bourg 97170, Guadeloupe
Location/Qualifiers
1. .863
/organism="Ehrlichia ruminantium"
/strain="Ball-3"
/db_xref="taxon:779"
1. .855
/notes="outer membrane protein; MAP1"
/codon_start=1
/transl_table=11
/product="major antigenic protein 1"
/protein_id="AAK27216.1"
/db_xref="GI:13448815"
/translation="MNCKIFITSLISLVSFLPGVSFSDVIQSDSPVSVYISAKY
MPTASHPGKMSIKEDSRDTRKVLGKMDGDKVTPSGNTNSIFTEKDYSPKVDNPPFL
FGAGAGYVMNGPRIEPEISYEVEDKNGKNDAMHVCALDTVSNYTAGATLSVM
VKGNLTDLSMLNACDIDMLDMPVSPYVCAGIGTDLVINATNPKLSVQKLGIS
YSINSETSIPIGGHFRVIGNEFRDITTSKVFNTSSGTSVSPGFASATLDVCHFI
EIGGRFVF"
291 a 127 c 165 g 279 t 1 others

```

```

Query Match      26.5%; Score 222.6; DB 1; Length 863;
Best Local Similarity 56.6%; Pred. No. 3.5e-30;
Matches 480; Conservative 0; Mismatches 350; Indels 18; Gaps 3;

Qy 1 ATGATTATAAGAAATCTTAGTAGAAGCGGTTAAATCTCATTAATGCTTACCA 60
Db 1 ATGATTGCAAGAAATTTTATCACAAGTACATAATATCATTTAGTGTCTTTTACCT 60

Qy 61 TATCAGTCTTTTGCAGATCTCTGATGTTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120
Db 61 GGTGTGTCCTTTTCTGATGTAATACAGGAAGACAGCAGCCAGT---AGGTAGTGTTC 117

Qy 121 ATTAGTCAAGATCAATCCAAAGTATATCAGTATATGATGATGATGATGATGATGATGAT 180
Db 118 ATTAGTCAAGATCAATCCAAAGTATATGATGATGATGATGATGATGATGATGATGAT 177

Qy 181 CCTA---TTAATGAACAAATTTCTCACTAAAGAAATTTTCGGACTAAAGAAAGATGGT 237
Db 178 TCTAGATATCAAGCAGATGTTTGGCTTTAAAGAAATTTGGGATGAGGATTTAAACACCA 237

Qy 238 GATATAACAAAAGAACAGATTTTACAAGAGTAGCTCCAGGCAATTTGATTTTCAAAAATAC 297
Db 238 TCGGTAACCAATTCATTTTACTGAAAAGATTTATCTTTTAAATATGATAATAT 297

Qy 298 TTAATATCAGGATTTTTCAGGAAGTATTTGGTTACTCTATGAGCGGACCAAGAAATAGAACTT 357
Db 298 CCATTTCTAGGTTTTCGCTGAGCAATTTGGGTACTCTNATGATGATGATGATGATGATGAT 357

Qy 358 GAAGCTGCATATCAACATTTTAAATCCAAAACACCGATACATGATGATGATGATGATGATGAT 417
Db 358 GAAATATCTCTATGAACTTTTGTATGTAAGAAACCCAGGTGGTAACTATAAAATGATGCC 417

```

QY 418 GAATCTATAAACAATTTTGCATTATCTCGTAAGATGCAATGGAAGATCAGCAATATGTA 477  
Db |||||  
QY 418 CACATGATTTGGCTCTAGATACAGATATCAAACTATATCGCGGAGCAATACATCTGTT 477  
Db |||||  
QY 478 GTACTTAAATAATGACGGCATAACTTTTATGTCATTGATGGTTAAATCTTGTCTATGACATT 537  
Db |||||  
QY 478 ATGGTAAATAACCGGAAATTTAACAGATATATCATTAATGTTAAATGCGGTGTTATGATATA 537  
Db |||||  
QY 538 ACAGCTGAAGAGTATCTTTGTCATCATATGATGTCAGGATATAGGAGCAGATCTTATC 597  
Db |||||  
QY 538 ATGCTTGATGGAATCCAGTGTCTCCATATGTTGTCAGGATTTGGCAGCTAGCTTAGTA 597  
Db |||||  
QY 598 ACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAATAGGTATAGTTAC 657  
Db |||||  
QY 598 TCAGTAATTAATGCTACAATCTAAATATCTTATCAAGGAAAGTTAGGTATAGCTAT 657  
Db |||||  
QY 658 CCTATCACCAAGAGTCTCTGCATTTTATGTTGGATACCTACATGCGGTTATTTGGTAAT 717  
Db |||||  
QY 658 TCTATCAATTCAGAAACATCTATATTTATTTGTTGGACACTTTTCATAGGTTATAGGAAAC 717  
Db |||||  
QY 718 AAATTTGAGAAGAT-----ACCTGTAATACTCTCTGTAGTATTAATGATGCT 765  
Db |||||  
QY 718 GAAATTAAGGATTAATCTACTTCTCAAGATTTTACTAATAGTAGTGGTACAGTAGTGT 777  
Db |||||  
QY 766 CCTCAACACCATCTGCTTCACTAACTCTTGACGTTGGATATCTTTGGCGGAGAAATTGA 825  
Db |||||  
QY 778 GTTACCCAGGTTTGCATCAGCAACCTTGATGTTGTCATTTGGCATAGAAATTGA 837  
Db |||||  
QY 826 ATGAGGTT 833  
Db |||||  
QY 838 GGAAGGTT 845

RESULT 24  
AF077733  
LOCUS  
DEFINITION  
Ehrlichia chaffeensis strain Jax outer membrane protein p28  
precursor, gene, complete cds.  
AF077733  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Ehrlichia chaffeensis.  
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
Anaplasmataceae; Ehrlichia.  
1 (bases 1 to 1243)  
Yu.X.J., McBride,J.W. and Walker,D.H.  
Genetic diversity of the 28-kilodalton outer membrane protein gene  
in human isolates of Ehrlichia chaffeensis  
J. Clin. Microbiol. 37 (4), 1137-1143 (1999)  
99175287  
10074538  
2 (bases 1 to 1243)  
Yu.X.-J. and Walker,D.H.  
Direct Submission  
Submitted (14-JUL-1998) Pathology, The University of Texas Medical  
Branch, 301 Univ. Blvd., Galveston, TX 77555, USA  
3 (bases 1 to 1243)  
Yu.X.-J. and Walker,D.H.  
Direct Submission  
Submitted (20-JUN-2001) Pathology, The University of Texas Medical  
Branch, 301 Univ. Blvd., Galveston, TX 77555, USA  
Sequence update by submitter  
On Jun 20, 2001 this sequence version replaced gi:3414968.  
Location/Qualifiers  
1. 1243  
/organism="Ehrlichia chaffeensis"  
/strain="Jax"  
/db\_xref="taxon:945"  
237. 1079  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane protein p28 precursor"

CDS

/protein\_id="AAC31546.1"  
/db\_xref="GI:3414969"  
/translation="MNYKKGVITTSALISLPGVSFSDPTSGSINGNFVISGKMP  
SASFGVFAKEERSTTGVFLKQDMGSAISHTTIFENITVSNFYKYNPFLGF  
AGATGYMDGPRIELEVSYETFDVKNQNSYKNEAHRYCALSHSSGTSMSAGDKFV  
FLKNEGILDVSMFLNACYDISEGIPEFYICVIGITDLVSMFEPATPKLSYQKGLG  
SYSISPTSVFIGHFHKVIGNEPRDIPAVIPNGSSLTGNHPAIVTLDVCHFGIELGG  
RFA"

BASE COUNT 396 a 197 c 233 g 417 t  
ORIGIN

Query Match 26.0%; Score 218.4; DB 1; Length 1243;  
Best Local Similarity 57.2%; Pred. No. 1.7e-29;  
Matches 487; Conservative 0; Mismatches 341; Indels 24; Gaps 4;

QY 1 ATGAATTATAGAAAAATCTTAGTAAGAGCGCGTTAATCTCATTAATGTCAATCTTACCA 60  
Db |||||  
QY 237 ATGAATTATAGAAAAATCTTAGTAAGAGCGCGTTAATCTCATTAATGTCAATCTTCTACCT 296  
Db |||||  
QY 61 TATCAGTCTTTTGCAGATCCTGTAGGTTCAAGAACTAATGATACAAAGAGGCTTCTAC 120  
Db |||||  
QY 297 GGAGTATCATTTTCTGACCCCAACAGGTAGTGGTATTACGGCAAT-----TTCTAT 347  
Db |||||  
QY 121 ATTAGTCAAAAGTACAATCCAAAGTATATCAACATTTTAGAAAAATTTCTGTCT---GAAGAA 177  
Db |||||  
QY 348 ATCAGTGAATAATATATGCCAAGCGCTTCGCATTTTGGGGTGTCTTCTGTAAAGAGAA 407  
Db |||||  
QY 178 ACTCCTATTAATGAACAAATCTCTCACTAAAAAGTTTTCGGACTAAAGAAAGATGCT 237  
Db |||||  
QY 408 AGAAGTACAAACAGCTGGAGTATTTGGATTGAAGCAAGATTGGGATGGCAGTGCATATCT 467  
Db |||||  
QY 238 GATATACAAAAAAAGACGATTTTACAAGAGTAGTCCAGGCATTTGATTTTCAAAATAAC 297  
Db |||||  
QY 468 CACACCCCCAGAAAATATATTCTGTTTCAAAATTTATTCGTTTAAATATGAANAATAC 527  
Db |||||  
QY 298 TTAATATCAGGATTTTTCAGGAAGTATTGGTTACTCTATGACGGACCAAGAAATAGAACTT 357  
Db |||||  
QY 528 CCATTTTATGATTTGACGAGGACTATTGGTTACTCAATGATGCCCCAAGAAATAGAGCTT 587  
Db |||||  
QY 358 GAAGTGCATATCAACAATTTAATCCAAAAACCGGATAACAATGATGATCTGTAATGCT 417  
Db |||||  
QY 588 GAAGTATCTTATGAGACATTCGATGTAAAAAATCAAGGTAAACAGTTACAAGAACGAAACG 647  
Db |||||  
QY 418 GAATACATAAACAATTTTGCATTTCTCGTAAAGATGCAAT-----GGAAGATCAG 468  
Db |||||  
QY 648 CATAGTATTTGCTTTTATCTCATCACAGTTTCAGGAACAAGCATGTCAGTGCAGGTAT 707  
Db |||||  
QY 469 CAATATGATGATCTTAAATAATGACGGCATAAATTTTATGTCATTTGATGTTTAAATCTTGC 528  
Db |||||  
QY 708 AAATTTGTTTCTTAAATAATGAAGGATTAATCTGACGATATCGTTTCATCTGCAACGATGC 767  
Db |||||  
QY 529 TATGACATTACAGCTGAAGGAGTATCTTCTGATCATATGATGTCAGGATATAGGACCA 588  
Db |||||  
QY 768 TATGATGTAATAAGTGAAGGAATACCTTTTCTCTTACATATGTTAGGTATCGGGACT 827  
Db |||||  
QY 589 GATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAATAGGT 648  
Db |||||  
QY 828 GATTTAGTATCCATGTTTGAAGCTACAATCTTAAATTTCTTACCAAGGAAAGTTAGGT 887  
Db |||||  
QY 649 ATTAGTTACCCCTATCACACCAAGAGTCTCTGCATTTTATTTGGTGGATATCTACCATGCGCT 708  
Db |||||  
QY 888 TTAAGCTACTCTATAAGCCCAAGAACTTCTGTGTTTATTTGGTGGACATTTTTCATAAGGTG 947  
Db |||||  
QY 709 ATTGTAATAAATTTTGAAGATACCTGTATAACTCTCTGTAGTATTAATGATGCTCTCT 768  
Db |||||  
QY 948 ATAGGAAACGAATTTAGAGATATTTCTGCTGTAAATACCTTAATGGATCAA---GTCTTACA 1004  
Db |||||  
QY 769 CAAACCATCTGCTTTCAGTAACTCTTGACGTTGGATCTTTTGGCGGAGAAATTTGGAATG 828  
Db |||||  
QY 1005 GGAACACACCTGCAATAGTAACACTGGACGATGTCACATTTGGTATAGAGCTTGGAGGA 1064  
Db |||||  
QY 829 AGGTTACCTTC 840  
Db |||||



```

Db      1065 AGTTTGGCTTC 1076

RESULT 25
AF077732
LOCUS      Ehrlichia chaffeensis strain 91HE17 outer membrane protein p28
DEFINITION precursor, gene, complete cds.
ACCESSION AF077732
VERSION    AF077732.2 GI:14488429
KEYWORDS   Ehrlichia chaffeensis.
SOURCE     Ehrlichia chaffeensis.
ORGANISM   Ehrlichia chaffeensis.
REFERENCE 1 (bases 1 to 1283)
AUTHORS   Yu, X.-J., McBride, J.W. and Walker, D.H.
TITLE     Genetic diversity of the 28-kilodalton outer membrane protein gene
          in human isolates of Ehrlichia chaffeensis
JOURNAL   J. Clin. Microbiol. 37 (4), 1137-1143 (1999)
MEDLINE   99175287
PUBMED    10074538
REFERENCE 2 (bases 1 to 1283)
AUTHORS   Yu, X.-J. and Walker, D.H.
TITLE     Direct Submission
JOURNAL   Submitted (14-JUL-1998) Pathology, The University of Texas Medical
          Branch, 301 Univ. Blvd., Galveston, TX 77555, USA
REFERENCE 3 (bases 1 to 1283)
AUTHORS   Yu, X.-J. and Walker, D.H.
TITLE     Direct Submission
JOURNAL   Submitted (19-JUN-2001) Pathology, The University of Texas Medical
          Branch, 301 Univ. Blvd., Galveston, TX 77555, USA
REMARK    Sequence update by submitter
COMMENT    On Jun 19, 2001 this sequence version replaced gi:3414966.
FEATURES   Location/Qualifiers
            source
              1..1283
              /organism="Ehrlichia chaffeensis"
              /strain="91HE17"
              /db_xref="taxon:945"
              236..1078
              /codon_start=1
              /transl_table=11
              /product="outer membrane protein p28 precursor"
              /protein_id="AAC31545.1"
              /db_xref="GI:3414967"
              /translation="MNYKVKVITGALISLISLPGVSFSDPTGSGINGNFYISKYMP
          SASHFGVFAKEERSTTAGVGLQDWDGSAISHTPENIFTVSNYSFKYENNPFLGF
          AGAIGYMDGPRILEVSYETPDYKQGNYSYKNEAHRYCALSHSSGTSMSAGDKFV
          FLKNEGLDVSFMLNACDVISEGPSPYICVIGIGLDVSMPEATPNKISYQKGLG
          SYSISPSTSVFPGHFKVIGNEFRDIPAVIPGSGSLTGNHNPATVLDVCHFGIELGG
          RFAF"

BASE COUNT 411 a 204 c 242 g 426 t

Query Match      26.0%; Score 218.4; DB 1; Length 1283;
Best Local Similarity 57.2%; Pred. No. 1.7e-29;
Matches 487; Conservative 0; Mismatches 341; Indels 24; Gaps 4;

Qy      1 ATGAATTAAGAAATTTCTAGTAGAGCGGTTATCTCATTAATGCTCAATCTTACCA 60
Db      236 ATGAATTAAGAAATTTCTATCAACAGTCGATGATATCACTAATCTTCTCTACCT 295
Qy      61 TATCAGTCTTTTGCAGATCCGTAGTTTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120
Db      296 GGAGTATCATTTTCTGACCAACAGGTAGTGTATTACGGCAAT-----TTCTAT 346
Qy      121 ATTAGTCAAGATCAATCCAGTATATCACTTTAGAAAATCTCTGCT---GAAGAA 177
Db      347 ATCAGTGAATAATATATGCCAAGCGCTTCGCATTTTGGGGTGTCTTCTGCTAAGGAAGAA 406
Qy      178 ACTCCTATTATGGAACAATCTCTCACTAAAAGTTTTCGCACTAAAGAAAGATGGT 237
Db      407 AGAAGTACACAGCTGGAGTATTTGGATTGAAGCAAGATTGGGATGGCAGTCAATATCT 466

```

```

Qy      238 GATATAACAAAAAGACGATTTTCAAGAGTAGCTCCAGGCATGTGATTTTCAAAATAAC 297
Db      467 CACACCAACCCAGAAAAATATATTCACCTGTTTCAAAATATTTCGTTTAAATATGAATAAC 526
Qy      298 TTATATACAGATTTTTCAGGAGTATTTGGTACTCTATGGACCGACCAAGATAGAACTT 357
Db      527 CCATTTTATAGATTTTGCAGGAGCTATTGGTTACTCAATGGATGGCCCAAGATAGAGCTT 586
Qy      358 GAAGCTGCATATCAACAATTTAATCCAAAAAACCCGATAACAATGATCATGATAATGGT 417
Db      587 GAAGTATCTTATGAGACATTCGATGTAAAAAATCAAGGTAAACAGTTACAGAAACGAACGG 646
Qy      418 GAATACTATAACAATTTTGGCATTATCTCGTAAAGATGCAAT-----GGAAGATCAG 468
Db      647 CATAGGTATTGTGCTTTTATCTCATCACAGTTCAGGAACAAGCATGTGCGAGTGCAGGTGAT 706
Qy      469 CAATATGTAGTACTTAAATAATGACGCATACATTTTATGTCATTTGATGGTTTAACTTGC 528
Db      707 AAATTTGTTTCTTAAAAAATGAAGGATTAATCTGACGTATCGTTTCATGCTGAACGCATGC 766
Qy      529 TATGACATTACAGCTGAAGGAGTATCTTTCGTACCATATGATGTGCGAGGTATAGGAGCA 588
Db      767 TATGATGTATAAGTGAAGGATACCTTTTCTCTTACATATGTGTAGGTATCGGACT 826
Qy      589 GATCTTATCACTATTTTAAAGACCTCAATCTAAATTTTGTCTTACCAAGGAAAAATAGGT 648
Db      827 GATTTAGTATCCATGTTTGAAGCTACAAATCTTAAATTTCTTACCAAGGAAAGTTAGGT 886
Qy      649 ATTAGTTACCTATACACCAAGAGTCTCTGCATTTATTTGTTGATGATACCTACCTGGCGTT 708
Db      887 TTAAGCTACTCTATAAGCCCAAGAACTTCTGTGTTTATTTGGTGGACATTTTTCATAAGGTG 946
Qy      709 ATTGGTAATAAATTTGAGAAGATACCTGTAATACTCTCTAGTATTAATTAATGATGCTCCT 768
Db      947 ATAGAAACGAAATTTAGAGATATTCCTGCTGTAATACCTAGTGGATCAA---GTCCTACA 1003
Qy      769 CAAACACATCTGCTTCAGTAACTCTTGACGTTTGGATACATTTTGGCGGAGAAATTTGGAATG 828
Db      1004 GGAAACCAACCTGCAATAGTAACACTGACGATGATGATGATGATGATGATGATGATGATGATG 1063
Qy      829 AGTTTCACTTC 840
Db      1064 AGTTTGGCTTC 1075

RESULT 26
AF393394
LOCUS      Ehrlichia chaffeensis strain V8 outer membrane protein p28 gene,
DEFINITION complete cds.
ACCESSION AF393394
VERSION    AF393394.1 GI:15991541
KEYWORDS   Ehrlichia chaffeensis.
SOURCE     Ehrlichia chaffeensis
ORGANISM   Ehrlichia chaffeensis
REFERENCE 1 (bases 1 to 1307)
AUTHORS   Long, S.W., Zhang, X.F., Qi, H., Standaert, S., Walker, D.H. and Yu, X.J.
TITLE     Antigenic variation of Ehrlichia chaffeensis resulting from
          differential expression of the 28-kilodalton protein gene family
JOURNAL   Infect. Immun. 70 (4), 1824-1831 (2002)
MEDLINE   21893092
PUBMED    11895944
REFERENCE 2 (bases 1 to 1307)
AUTHORS   Yu, X.-J. and Zhang, X.-F.
TITLE     Direct Submission
JOURNAL   Submitted (20-JUN-2001) Pathology, University of Texas Medical
          Branch, 301 Univ. Blvd., Galveston, TX 77555-0609, USA
FEATURES   Location/Qualifiers
            source
              1..1307
              /organism="Ehrlichia chaffeensis"

```

```
/strain="V8"
/db_xref="taxon:945"
236..1078
/notes="p28-19; 28 kDa outer membrane protein"
/codon_start=1
/transl_table=11
/product="outer membrane protein p28"
/protein_id="RAL12924.1"
/db_xref="GI:15991542"
/translation="MNYKKVFTSALISLSLPGVFSDFDPTSGINGNFYISGKMP
SASHFGVSAKEERSTTAGVFLKQDWDGSAISHTPENIFVSNYSFKYNNPFLGF
AGATGYSDMDGPRIELEVSYETFDVKNQNSYKNEAHRYCALSHSSGTSMSAGDKFV
FLKNEGLDVSFMLNACVDVISEGIPRSPYICVGTDLVSMFEATNPKI SYOGKIGL
SYSISPSTSVFIGHFHVKVIGNEPRDIPAVIPSSSLTGNHPAIVLTDVCHFGIELG
RFAP"
BASE COUNT      420 a   212 c   245 g   430 t
ORIGIN
Query Match      26.0%; Score 218.4; DB 1; Length 1307;
Best Local Similarity 57.2%; Pred. No. 1.7e-29;
Matches 487; Conservative 0; Mismatches 341; Indels 24; Gaps 4;

Qy 1 ATGAATTATAGAAAATTTCTAGTAAGAAGCGGTTTAATCTCAATTAATGCTCAATCTTACCA 60
Db 236 ATGAATTACAAAAGTTTTCATAACAAGTCATTGATATCACTAATATCTTCTCTACCT 295
Qy 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAGAACTAATGATAACAAGAAGGCTTCTAC 120
Db 296 GGAGTATCATTTTCTGACCCACAGGTAGTGTATTAAACGGCAAT-----TTCTAT 346
Qy 121 ATTAGTCAAAAGTACAATCCAAGTATATCACACTTTAGAAAATTTCTCTGCT---GAAGAA 177
Db 347 ATCAGTGAAATATATGCCAAGCCCTCGCATTTTGGGGTGTTCCTGCTTAAGGAAGAA 406
Qy 178 ACTCCTATTAAATGGAACAAATTTCTCTACTAAAAAGTTTTCGGACTAAAGAAAGATGGT 237
Db 407 AGAAGTACAACAGCTGGAGTATTGGATTGAAGCAAGATTGGAGTGGCAGTCAATATCT 466
Qy 238 GATATAACAAAAGAGAGATTTTACAGAGTAGTCTCCAGGATTTGATTTTCAAAATAC 297
Db 467 CACACCACCCCAAGAAATATATTCATCTGTTTCAAAATTTTCGTTTAAATATGAAAATAAC 526
Qy 298 TTAATATCAGGATTTTCAGGAAGTATTGGTTACTCTATGGAGCGGACCAGAAATAGAACTT 357
Db 527 CCATTTTATAGATTTCGAGGAGCTATTGGTTACTCAATGGATGGCCCAAGAAATAGAGTT 586
Qy 358 GAAGCTGCATATCAACAAATTTAATCCAAAAACACCGATAACATGATCTGATAATGGT 417
Db 587 GAAGTATCTTATGAGACATTCGATGTAAAAAATCAAGGTAAACAGTTACAAGAACCAAGCG 646
Qy 418 GAATACTATAACATTTTGCATTATCTCGTAAAGATGCAT-----GGAAAGATCAG 468
Db 647 CATAGGTATTGTGCTTTATCTCATCACAGTTCAGGAACAAGCATGTCGAGTGCAGGTGAT 706
Qy 469 CAATATGTAGTACTTAAAAATGACGGCATACTTTTATGTCATTCATGTTTAATCTTGC 528
Db 707 AAATTTGTTTCTAAAAAATGAAGATTACTTGACGTATCTTCACTGTCATGACCCATGCG 766
Qy 529 TATGACATTACAGCTGAAGGAGTATCTTTCGTAACCATATGCAATGTCGAGGTATAGAGCA 588
Db 767 TATGATGTAATAAGTGAAGGAATACCTTTTCTCCTTACATATGTTGTAGGTATCGGACT 826
Qy 589 GATCTTATCACTATTTTAAAGACCTCAATCTAAAATTTGCTTACCAAGGAATAATAGGT 648
Db 827 GATTTAGTATCCATGTTTGAAGCTACAAATCTTAAATTTCTTACCAAGGAAGTATAGGT 886
Qy 649 ATTAGTACCCTTATCACACCAAGTCTCTGCAATTTATTTGGTGGATACCTACCATCGCTT 708
Db 887 TTAAGTACTCTATAAGCCCAAGAACTCTGTGTTTATTTGGTGGACATTTTCAATAGGTG 946
Qy 709 ATTGTAATAAATTTGAGAAGATACCTGTAATACTCTCTGTAGTATTAAATGATCTCCT 768
Db 947 ATAGGAACGAATTTAGAGATATTCTCTGCTGTAATACCTAGTGGATCAA---GTCTTACA 1003
```

```
769 CAACACCATCTCTCAGTAACCTTTGACGTTGGATACCTTTGGCGGAGAAATTTGGAATG 828
1004 GGAAACCAACCCCTGCAATAGTAACACACTGGACGATGCCCATTGTTGATATAGACTTTGGAGGA 1063
829 AGGTTTCACCTTC 840
1064 AGGTTTGTCTTC 1075

RESULT 27
AP393390 1309 bp DNA linear BCT 27-MAR-2002
LOCUS
DEFINITION Ehrlichia chaffeensis strain V4 outer membrane protein p28 gene,
complete cds.
ACCESSION AF393390
VERSION AF393390.1 GI:15991533
KEYWORDS
SOURCE
ORGANISM
Ehrlichia chaffeensis.
Ehrlichia chaffeensis
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE
1 (bases 1 to 1309)
Long, S.W., Zhang, X.F., Qi, H., Standaert, S., Walker, D.H. and Yu, X.J.
Antigenic variation of Ehrlichia chaffeensis resulting from
differential expression of the 28-kilodalton protein gene family
Infect. Immun. 70 (4), 1824-1831 (2002)
JOURNAL
MEDLINE 21893092
PUBMED 11895944
REFERENCE
2 (bases 1 to 1309)
Yu, X.-J., Zhang, X.-F. and Walker, D.H.
Direct Submission
AUTHORS
TITLE
JOURNAL
SUBMITTED (20-JUN-2001) Pathology, University of Texas Medical
Branch, 301 Univ. Blvd., Galveston, TX 77555-0609, USA
Location/Qualifiers
FEATURES
source
1..1309
/organism="Ehrlichia chaffeensis"
/strain="V4"
/db_xref="taxon:945"
238..1080
/notes="p28-19; 28 kDa outer membrane protein"
/codon_start=1
/transl_table=11
/product="outer membrane protein p28"
/protein_id="RAL12920.1"
/db_xref="GI:15991534"
/translation="MNYKKVFTSALISLSLPGVFSDFDPTSGINGNFYISGKMP
SASHFGVSAKEERSTTAGVFLKQDWDGSAISHTPENIFVSNYSFKYNNPFLGF
AGATGYSDMDGPRIELEVSYETFDVKNQNSYKNEAHRYCALSHSSGTSMSAGDKFV
FLKNEGLDVSFMLNACVDVISEGIPRSPYICVGTDLVSMFEATNPKI SYOGKIGL
SYSISPSTSVFIGHFHVKVIGNEPRDIPAVIPSSSLTGNHPAIVLTDVCHFGIELG
RFAP"
BASE COUNT      420 a   212 c   247 g   430 t
ORIGIN
Query Match      26.0%; Score 218.4; DB 1; Length 1309;
Best Local Similarity 57.2%; Pred. No. 1.7e-29;
Matches 487; Conservative 0; Mismatches 341; Indels 24; Gaps 4;

Qy 1 ATGAATTATAGAAAATTTCTAGTAAGAAGCGGTTTAATCTCAATTAATGCTCAATCTTACCA 60
Db 238 ATGAATTACAAAAGTTTTCATAACAAGTCATTGATATCACTAATATCTTCTCTACCT 297
Qy 61 TATCAGTCTTTTGCAGATCCTGTAGTTCAGAACTAATGATAACAAGAAGGCTTCTAC 120
Db 298 GGAGTATCATTTTCTGACCCACAGGTAGTGTATTAAACGGCAAT-----TTCTAT 348
Qy 121 ATTAGTCAAAAGTACAATCCAAGTATATCACACTTTAGAAAATTTCTCTGCT---GAAGAA 177
Db 349 ATCAGTGAATAATATATGCCAAGCGCTTCGCATTTTGGGGTGTTCCTGCTTAAGGAAGAA 408
Qy 178 ACTCCTATTATGGAACAAATTTCTCTCACTAAAAAGTTTTCGGACTAAAGAAAGATGGT 237
```

```

Db 409 AGAAGTCAACAGCTGGAGTATTTGGATTGAAGCAAGATTGGATGGCAGTGCATATCT 468
Qy 238 GATATAACAAAAAGACGATTTTACAAAGTAGTCCAGGCAATGATTTTCAAAATAAC 297
Db 469 CACACCCAGGAGGATATATCTACTGTTTCAAAATATTCTGTTTAAATATGAAATAAC 528
Qy 298 TTAATATCAGGATTTTCAAGGAGTATTTGGTTACTCTATGAGCGGACCAAGATAGACTT 357
Db 529 CCATTTTAGGATTTGCGAGGAGTATTTGGTTACTCAATGGATGGCCCAAGATAGAGCTT 588
Qy 358 GAAGCTGCATATCAACATTTTAAATCCAAAAACACCGGATAACAATGATGATGATAATGGT 417
Db 589 GAAGTATCTTATGACATCTCGATGTAATAAATCAAGTACAGTTACAGAAACGAGCG 648
Qy 418 GAATACTATAACATTTTGCATATCTCGTAAAGATGCAAT-----GGAAGATCAG 468
Db 649 CATAGGTATTTGCTTTATCTCATCACAGTTTCAAGAACAGCATGTCGAGTGCAGGTGAT 708
Qy 469 CAATATGTAGTACTTAAATGACGGCATACTTTTATGTCATTCATGCTTAACTTGC 528
Db 709 AATTTGTTTTTCTAAAAAATGAAGGATTTCTGACGATGCTGCTGACGCGCATGC 768
Qy 529 TATGACATTACAGCTGAAGGAGTATCTTCTGACCATATGATGTCAGGATATAGGACA 588
Db 769 TATGATTAATAGTGAAGGATACCTTTTCTCTTACATATGTTAGGATCGGACT 828
Qy 589 GATCTTATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGGAAAAATAGGT 648
Db 829 GATTTAGTATCCATGTTTGAAGTACAAATCTTAAATTTCTTACCAAGGAAATAGGT 888
Qy 649 ATTAGTACCTATCACACGAGAGTCTCTGATTTTATGTTGATGATACCATCGGCTT 708
Db 889 TTAAGCTACTCTAAGGCCAGAACTTCTGTGTTTATTTGTTGACATTTTTCATAAGGTG 948
Qy 709 ATTGTAATAAATTGAGAAGATACCTGTAATACTCTGTAATTAATGATGCTCCT 768
Db 949 ATAGAAACGATTTAGAGATTTCTGCTGTAATACCTAGTGGATCAA----GTCCTACA 1005
Qy 769 CAACACCATCTGTTTCACTAGTAACTCTTGACGTTGGATCTTTGGCGGAGAAATGGAATG 828
Db 1006 GGAACACCCCTGCAATAGTAACACTGGAAGTATGCCACTTTTGGTATAGAGCTTGGAGGA 1065
Qy 829 AGGTTCACTTC 840
Db 1066 AGGTTGCTTTC 1077

```

```

RESULT 28
AF393391
LOCUS   Ehrlichia chaffeensis strain V5 outer membrane protein p28 gene,
DEFINITION complete cds.
ACCESSION AF393391
VERSION   AF393391.1 GI:15991535
KEYWORDS
SOURCE    Ehrlichia chaffeensis.
ORGANISM  Ehrlichia chaffeensis
          Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
          Anaplasmataceae; Ehrlichia.
REFERENCE 1 (bases 1 to 1189)
AUTHORS   Long, S.W., Zhang, X.F., Qi, H., Standaert, S., Walker, D.H. and Yu, X.J.
TITLE     Antigenic variation of Ehrlichia chaffeensis resulting from
          differential expression of the 28-kilodalton protein gene family
          Infect. Immun. 70 (4), 1824-1831 (2002)
JOURNAL   21893092
MEDLINE   11895944
PUBMED
REFERENCE 2 (bases 1 to 1189)
AUTHORS   Yu, X.-J. and Zhang, X.-F.
TITLE     Direct Submission
JOURNAL   Submitted (20-JUN-2001) Pathology, University of Texas Medical
          Branch, 301 Univ. Blvd., Galveston, TX 77555-0609, USA
FEATURES
          Location/Qualifiers
          1..1189

```

```

/organism="Ehrlichia chaffeensis"
/strain="V5"
/db_xref="taxon:945"
172..1014
/notes="p28-19; 28 kDa outer membrane protein"
/codon_start=1
/transl_table=11
/product="outer membrane protein p28"
/protein_id="AALI12921.1"
/db_xref="GI:15991536"
/translation="MNYKVFILKSLISSLPVSPDPAAGINGNFIYRGKYP-
SAGHPGVSAKERSSTAGVGLKQDWDGSAISHTTBNIFTVSNFKEVTPPLGF
AGAICYMDGPRIELEVSYETFDVKNGCNSYKNAHRYCALSHSSGTSMSAGDKFV
FLKNEGGLDVSFMLNACYDISEGIPSPYICVIGITDLSFMFATNPKISYQKGLG
SYSISPSTSVFIGGHPKHVINEFRDIPVIPSGLSTGNHPIATVITLVDVCHGIELG
RFAP"
BASE COUNT 380 a 192 c 216 g 401 t
ORIGIN
Query Match 25.6%; Score 215.2; DB 1; Length 1189;
Best Local Similarity 56.9%; Pred. No. 6.5e-29;
Matches 485; Conservative 0; Mismatches 343; Indels 24; Gaps 4;
Qy 1 ATGAATTTATAGAAATTTCTAGTAAGAGCGCGTTAATCTCAATTAATGCTCAATCTTACCA 60
Db 172 ATGAATTTACAAAAGTTTTTCATAAACAAGTGCATTGATATCATCAATATCTTCTACCT 231
Qy 61 TATCAGTCTTTTTCAGAGATCTGTAGTTCAGAACTAATGATAACAACAAAGAGGCTTCTAC 120
Db 232 GGAGTATCATTTTCTGACCAGAGGTAGGGTATTAACGGCAAT-----TTCTAT 282
Qy 121 ATTAGTCAAGTACAATCAAGTATATCACATTTTAGAAAATTTCTGTCT---GAAGAA 177
Db 283 ATCAGGGGAAAATATATGCCCAGCGCTTCGCAATTTTGGGGTGTCTTCTGCTAAGGAAGA 342
Qy 178 ACTCTATTATGGAACAAATTTCTCACTAAAAAGTTTTCGACTTAAGAAAGATGGT 237
Db 343 AGAAGTACAACAGCTGGAGTATTTTGAAGTGAAGCAAGATTGGGATGGCAGTGCATATATCT 402
Qy 238 GATATAACAAAAAGACGATTTTACAAAGTAGTCTCCAGGCAATTTTTCAAAAATAAC 297
Db 403 CACACACCCAGGAAATATATTCATCTGTTTCAATATTTCTGTTTAAATATGAAATACC 462
Qy 298 TTAATATCAGGATTTTTCAGGAGTATTTGGTTACTCTATGACGCGACCAAGAAATAGAACTT 357
Db 463 CCATTTTGGATTTTTCAGGAGCTATTTGGTTACTCAATGGATGGCCCAAGAAATAGAGCTT 522
Qy 358 GAAGCTGCATATCAACATTTTATCCAAAAACACCGATACAATGATGATGATAATGGT 417
Db 523 GAAGTATCTTATGAGACATTTGATGTAATAAATCAAGGTAAACAGTTACAGAACGAAAGCG 582
Qy 418 GAATACTATAACATTTTGCATATCTCTCATCAGTTTCAGAAACAGCATGTCGAGTGCAGGTGAT 642
Db 583 CATAGGTATTTGCTTTATCTCATCAGTTTCAGAAACAGCATGTCGAGTGCAGGTGAT 642
Qy 469 CAATATGTAGTACTTAAAAATGACGGCAATCTTTTATGTCATTTGATGGTGGTAACTTTC 528
Db 643 AAATTTGTTTTTCTAAAAAATGAAGGATTTACTTGACGATCTGTTTCATGCTGAACGATGC 702
Qy 529 TATGACATTACAGCTGAAGGAGTATCTTTCGTACCATATGATGATGATGATGATGAGGACA 588
Db 703 TATGATGTAATAAGTGAAGGAAATCCCTTTTCTCTTACATATGTTGATGATGATGAGGACT 762
Qy 589 GATCTTATCACTATTTTAAAGACCTCAATCTAAATTTTGTCTTACCAAGGAAAAATAGGT 648
Db 763 GATTTAGTATCCATGTTTGAAGCTACAATCTTAAATTTTCTTACCAAGGAAATAGGT 822
Qy 649 ATTAGTTACCTATCACACGAGAGTCTCTGCAATTTTATGTTGGTGGATCTACTACCATGGCTT 708
Db 823 TTAAGCTACTCTAAGCCCAAGAACTTCTGTGTTTATTTGTTGGGACATTTTTCATAAGGTG 882
Qy 709 ATTGGTAAATAATTTGAGAGATACCTGTAATAACTCTCTGTAGTATTAATGATGCTCCT 768

```

```

Db      883 ATAGAAACGAATTTAGAGATATCTCTGCTAATACCTAGTGGATCAA---GTCTTACA 939
Qy      769 CAACCAACATCTGCTTCACTAATCTTTCAGCTTGGATATCTTTGGCGGAGAAATTCGAATG 828
Db      940 GGAACCAACCTCTGCAATAGTACACTGGACGATGCCACTTTGGTATATAGAGCTTGGAGGA 999
Qy      829 AGGTTCCACTTC 840
Db      1000 AGGTTGCTTTC 1011

RESULT 29
AF393395
LOCUS   AF393395
DEFINITION   Ehrlichia chaffeensis strain V9 outer membrane protein p28 gene,
complete cds.
ACCESSION   AF393395
VERSION     AF393395.1 GI:15991543
KEYWORDS    '
SOURCE      Ehrlichia chaffeensis.
ORGANISM    Ehrlichia chaffeensis
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE   1 (bases 1 to 1282)
AUTHORS    Long, S.W., Zhang, X.F., Qi, H., Standaert, S., Walker, D.H. and Yu, X.J.
TITLE      Antigenic variation of Ehrlichia chaffeensis resulting from
different expression of the 28-kilodalton protein gene family
JOURNAL    Infect. Immun. 70 (4), 1824-1831 (2002)
MEDLINE    21893092
PUBMED     11895944
REFERENCE   2 (bases 1 to 1282)
AUTHORS    Yu, X.-J. and Zhang, X.-F.
TITLE      Direct Submission
JOURNAL    Submitted (20-JUN-2001) Pathology, University of Texas Medical
Branch, 301 Univ. Blvd., Galveston, TX 77555-0609, USA
FEATURES   source
            1..1282
            /organism="Ehrlichia chaffeensis"
            /strain="V9"
            /db_xref="taxon:945"
            214..1044
            /note="p28-19; 28 kDa outer membrane protein"
            /codon_start=1
            /transl_table=11
            /product="outer membrane protein p28"
            /protein_id="AAL12925.1"
            /db_xref="GI:15991544"
            /translation="MNYKKVFTTSLISLPGVSPDPTGSGINGNFVISGKMP
            SASHFVPSAKEERTTAGVGLKODWDSALPHTHSNTPTVSNYSFKYNNPLGF
            AGATGYSNDGPRIELEVSVEYFDVKNQNNYKNEAHRVCALSRQNTIADNTFFLKN
            EGLDISFLNACYDIVISEGIPFSPYICAGIGTDLSWFMFEVTPNPKISYOGKLGUSYI
            SPETSVFVGGHFHKVMGNFEFRDPAIIPSGSSLTGNHPAIVLTDVCHFGIELGGRFAP
            "

BASE COUNT  425 a 218 c 212 g 427 t
ORIGIN

Query Match 25.4%; Score 213.6; DB 1; Length 1282;
Best Local Similarity 57.7%; Pred. No. 1.2e-28;
Matches 488; Conservative 0; Mismatches 334; Indels 24; Gaps 5;

Qy      1 ATGATTATAGAAATTTCTAGTAGAGAGCGGTTAACTCAATCTCAATATGCTCAATCTTACCA 60
Db      214 ATGAATTACAAAAGTTTTCATACAAGTGCATTGATATCACTAATATCTTCTCTACCT 273
Qy      61 TATCAGTCTTTTGCAGATCTCTGAGTTTCAAGAACTAATGATATACAAAGAGCGTTCTTAC 120
Db      274 GGAGTATCATTTTCGCCACACAGTAGTGTGTTAATACGGCAAT-----TTCTAT 324
Qy      121 ATTAGTGAAGTACAATCAAGTATATACACTTTAGAAAATTTCTTGCT---GAAGAA 177
Db      325 ATCAGTGAAGAAATATATGCAAGCGCTTGGCAATTTGGGGTGTGTTTCTGCTAAGGAAGAA 384
Qy      178 ACTCCTATTATGGAACAAATTTCTCTCACTAAAAAAGTTTTTCGGACTAAAGAAAGATGCT 237

```

```

Db      385 AGAATACACAGCTGGAGTATTTGGATTGAGCAAGATTGGGATGCGAGTGCATACCT 444
Qy      238 GATATAACAAAAAGACGATTTTACAAGAGTAGCTCCAGGCATTTGATTTTCAAAAATAC 297
Db      445 CACACTCATAGTAATACCTCCATTTACTGTTTCAAACTACTCATTTAAATATGAAAAATAT 504
Qy      298 TTAATATCAGGATTTTTCAGGAGTATTTGGTTACTCTATGGACGACCAAGAAATAGACTT 357
Db      505 CCATTTCTAGGTTTTCAGGAGGACTTTTGGCTACTCAATGGATGGTCCAAAGAAATAGAGCTT 564
Qy      358 GAAGCTGCATATCAACAATTTAATCCAAAAAACACCGATACCAATGATCTACTGATAATGGT 417
Db      565 GAAGTATCTTATGAGACATTTGATGTTAAAAATCAAGTAACAA-----CTATAAGAAC 618
Qy      418 GAATACTATAAACATTTTGCATTAATCTCGTAAAGATGCAAT---GGAAGATCAGCAATAT 474
Db      619 GAAGCTCATAGTACTGTGCTTTATCCGTCAGGCCACAATAGCAGCAGATAATACATTT 678
Qy      475 GTAGTACTTAAAAATGAGGCAATACCTTTTATGTCATTGATGGTTAATACTTGTCTATGAC 534
Db      679 GTTTTCTTAAAAAATGAAGGGCTACTTGACATATCATTTATGCTGAACGCAATGCTATGAT 738
Qy      535 ATTACAGCTGAAGGAGTATCTTTCGTACCATATGCATGTGCAGGTATAGGAGCAGATCTT 594
Db      739 GTATAAGTGAAGAAATACCTTTTCTCTTACATATGTGCAGGTATTTGGTACTGATTTA 798
Qy      595 ATCACTATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGAAAAAATAGGTATTAGT 654
Db      799 GTATCCATGTTTGAAGTTTACAAACCTTAAAAATTTCTTACCAAGGAAAGTTAGGTTTAAGC 858
Qy      655 TACCCTATCACACGAGAGCTCTGCATTTTATTTGGTGTACTACTACCGGCTTATTTGGT 714
Db      859 TACTCTATAAGCCCGAGAAACTTCTGCTTTGTCGGCGGACATTTCCATAAGGTAAATGGGA 918
Qy      715 AATAAAATTGAGAAGATACCTGTAAATACTCCTCTAGTATTAAATGATGCTCTCTCAAAACC 774
Db      919 AACGAATTCAGAGATATTCCTGCTATATACCTAGTGGATCAAGCTTTACAGGGAATCAC 978
Qy      775 ACATCTGCTTCAGTAACTCTTTCAGCTTGGATPACTTTTGGCGGAGAAATTTGGAATGAGGTTTC 834
Db      979 ---CCTGCAATAGTAAACACTAGATGTATGCCACTTTGCTATAGAACTTTGGAGGAAGATTT 1035
Qy      835 ACCTTC 840
Db      1036 GCTTTC 1041

RESULT 30
AF077734
LOCUS   AF077734
DEFINITION   Ehrlichia chaffeensis strain Sapulpa outer membrane protein p28
precursor, gene, complete cds.
ACCESSION   AF077734
VERSION     AF077734.2 GI:14495546
KEYWORDS    '
SOURCE      Ehrlichia chaffeensis.
ORGANISM    Ehrlichia chaffeensis
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE   1 (bases 1 to 1312)
AUTHORS    Yu, X.J., McBride, J.W. and Walker, D.H.
TITLE      Genetic diversity of the 28-kilodalton outer membrane protein gene
in human isolates of Ehrlichia chaffeensis
JOURNAL    J. Clin. Microbiol. 37 (4), 1137-1143 (1999)
MEDLINE    99175287
PUBMED     10074538
REFERENCE   2 (bases 1 to 1312)
AUTHORS    Yu, X.-J. and Walker, D.H.
TITLE      Direct Submission
JOURNAL    Submitted (14-JUL-1998) Pathology, The University of Texas Medical
Branch, 301 Univ. Blvd., Galveston, TX 77555, USA
REFERENCE   3 (bases 1 to 1312)

```

QY	655	TACCCTATCACCCAGAAAGTCTCTGCATTTATTGGTGGATACTACCATGGCGTTATTGGT	714
Db	888	TACTCTATAAGCCAGAAACTCTGTCTTTGTGCGGGACATTTCCATAAGGTAAATGGGA	947
QY	715	AATAAATTTGAGAAGATACCTGTATAACTCTCTGTAGTATTAAATGATGCTCTCAACC	774
Db	948	AACGAATTCAGAGATATTCCTGCTATAATACCTAGTGGATCAAGTCTTACAGGAATCAC	1007
QY	775	ACATCTGCTTCAGTAACTCTTGACGTTGGATACTTTGGCGGAGAAATTCGAAATGAGGTTTC	834
Db	1008	---CCTGCAATAGTAACACTAGATGTATGCCACTTTTGGTATGAACCTTGGAGGAAGATTT	1064
QY	835	ACCTTC 840	
Db	1065	GCTTTC 1070	
RESULT 31			
LOCUS	AX042312	861 bp	DNA
DEFINITION	Sequence 8 from Patent WO0065063.		linear
ACCESSION	AX042312		PAT 23-NOV-2000
VERSION	AX042312.1	GI:11340972	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
BASE COUNT			
ORIGIN			
Query Match	25.3%;	Score 212.8;	DB 6;
Best Local Similarity	57.3%;	Pred. No. 1.9e-28;	Length 861;
Matches	492;	Conservative 0;	Mismatches 342;
			Indels 24;
			Gaps 5;
QY	1	ATGAATATTAAGAAAATCTAGTAAAGCGGTTAATCTCAATTAATGTCAATCTTTACCA	60
Db	3	ATGAATCTGCGAAAATTTTTTATAACAATCGCATTAACATTACTAATGTCTTTACCT	62
QY	61	TATCAGTCTTTTGGAGATCCTGTAGTTCAGAACTAATGATGAACAAGAGCTTCTAC	120
Db	63	GGAAATATCACTTTCTGATCCAGTACAG-----GATGACAACATAGTGGTAATTTCTAC	116
QY	121	ATTAGTGCAGGTACAAATCCAGTATATCACACTTTTAGAAAATTTCTCTGCT---GAAGAA	177
Db	117	ATCAGTGGAAAGTATATGTCACAGCGCTTCGCAATTTTGGAGTTTTTTCTGCCAAGGAAGAA	176
QY	178	ACTCTATTAAATGGAAACAAATTTCTCTCACTAAAAAAGTTTTTCGGACTAAAGAAAGATGGT	237
Db	177	AGAAATACAACAGTTGGAGTATTTGGAAATAGACAGATTTGGATAGATGTGTAATATCT	236
QY	238	GATATAACAAAAAGACGATTTTACAGAGTAGTCCAGGCAATTGATTTTCAAAATAAC	297
Db	237	AGAACCCATTTAAGCGATATATTCACCGTTCCAAATTTATTCATTTAAGTATGAAATAAT	296
QY	298	TTAATATCAGGATTTTCAGAGTATTTGGTTACTCTATGACGGACCAAGATAGACATT	357
Db	297	CTATTTTCAGGATTTGACAGGACTATTGGCTACTCAATGGATGGCCCAAGATAGAGCTT	356
QY	358	GAAGCTGCATATCAACAAATTTTAATCCAAAAAACCCGATAACAAATGATCTGATATG--	415



Qy

```

Db      3  ATGAATTCGAAATAATTTTATACAACTACATAGTATCGTAATGCTCTTACT 62
Qy      61  TATCAGTCTTTGAGATCTGTAGGTTCAAGAACTAATGATAACAAAGAGGCTTTAC 120
Db      63  GGAATATCATTTCTGATCGAGTACAGAACGACAAATGTTGGTGAAT-----TTCTAT 116
Qy      121  ATTAGTCAAGATCAATCCAGTATATACACTTTTGAATAATCTCTGCTGAA---GAA 177
Db      117  ATCAGTGGGAATATGTACCAAGTGTTCACATTTTGGCGTATCTCTGCTAAACAGAA 176
Qy      178  ACTCCTATTAAATGGAACAAATCTCTCACTAAATAAGTTTTCGGACTAAAGAAAGATGCT 237
Db      177  AGAATAACAACATCGGAGTATTGGATTAAGCAAGATTGGATGGCAGCACATATCT 236
Qy      238  GATATAACAAAAGACGATTTTACAAGAGTAGTCCAGGCAATGATTTTCAAAATAAC 297
Db      237  AAAAATCTCCAGAAAATACATTTAAACGTTCCAAATATTATTTAAATATGAAAATAAT 296
Qy      298  TTAATATCAGGATTTTCAGGAGTATTGGTTACTCTATGACGACCAAGATAGAACTT 357
Db      297  CCATTTCTAGGTTTTCAGGAGCTGTTGGTTATTTAATGAATGGTCCAGAAATAGAGTTA 356
Qy      358  GAAGCTGCATATCAACAATTTAATCCAAAACACCGGATAACAATGATCTGATAATGCT 417
Db      357  GAATATGCTATGAACAATTTGATGTGAATAACAGGTAATTAATTAAGAACGATGCT 416
Qy      418  GAATACTATAACATTTTGCATTAATCTCGT-----AAAGATGCAATGGAAGATCAGAA 471
Db      417  CACAAATATTATGCTTTAAACCCATAACAGTGGGGGAAAGCTAAGCAATGCAGGTGATAAG 476
Qy      472  TATGTAGTACTTAAATAATGACGGCACTATTTATGTCATGATGGTTAATACTTGGCTAT 531
Db      477  TTTGTTTTCATAAATAATGAAGGACTTCTGTATATATCATCTATGTTGAATGCATGCTAT 536
Qy      532  GACATTACAGCTGAAGGAGTATCTTTCGTACCATATGATGTCAGGTATAGGACAGAT 591
Db      537  GATGTAAATGAGGAATACCTTCTCTCTTACATATGTCAGGTGTTGGTACTGAT 596
Qy      592  CTTATCACTATTTTAAAGACTCAATCTAAATTTGCTTACCAAGGAAAATAGGTATT 651
Db      597  TTAATATCCATGTTTGAAGCTATAACCTTAAATTTCTTATCAAGGAAAGTTAGGTTTG 656
Qy      652  AGTTACCTATCACACGAGTCTCTGCAATTTATTTGGTGGATACATACATGCGGTTATT 711
Db      657  AGTTACTCCATAAGCCGAGAGCTTCTGTTTGTGTTGGACATTTTTCATAAGGTGATA 716
Qy      712  GGTATAAATTTGAGAAGATACCTGTAATACTCTGTAGTATTAAATGATGCTCCTCAA 771
Db      717  GGGATGAATTCAGAGATATTCCTGCTATGATACCCAGTACCTCACTCTCACAGGTAT 776
Qy      772  ACCATCTGTTTCAGTAACTCTTGACGTTGGATATCTTTGGCGGAGAAATGGAAATGAG 831
Db      777  CAC---TTTACTATAGTAACACTAAGTGTATGCCACTTTGGAGTGGAACTTTGGAGGAGG 833
Qy      832  TTCACCTT 839
Db      834  TTTAACTT 841

```

```

RESULT 35
LOCUS   CRU50830
DEFINITION Cowdria ruminantium Antigua major antigenic protein 1 (map1) gene, complete cds.
ACCESSION U50830
VERSION   U50830.1
KEYWORDS  GI:1418249
SOURCE   heartwater rickettsia strain=Antigua.
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Anaplasmataceae; Ehrlichia.
REFERENCE 1 (bases 1 to 1263)

```

```

AUTHORS Reddy,G.R., Sulsona,C.R., Harrison,R.H., Mahan,S.M., Burridge,M.J.
and Barbet,A.F.
TITLE Sequence heterogeneity of the major antigenic protein 1 genes from Cowdria ruminantium isolates from different geographical areas
JOURNAL Clin. Diagn. Lab. Immunol. 3 (4), 417-422 (1996)
MEDLINE 96400830
PUBMED 8807206
REFERENCE 2 (bases 1 to 1263)
AUTHORS Reddy,G.R.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-1996) Roman G. Reddy, Pathobiology, University of Florida, 471 Mowry Rd., Gainesville, FL 32611-0880, USA
FEATURES
source
1..1263
/organism="Ehrlichia ruminantium"
/strain="Antigua"
/db_xref="taxon:779"
397..1242
/gene="map1"
397..1242
/gene="map1"
/translation="MNCCKIFITSLISLVFLPGVSFSDVIOEDSPFSGSVYISAKY
MPTAGHFGKMSIKEDSRDVKVFLGKDDGVKTSSTNTFTKDYDFKYENNPFLGF
AGTAHYGMNGPRIEFEISYETFDVFNPGNVDKNDAHMYCALDITATSGGAAATSVMV
KNEMLTDISILMNACYDILMDGMPVSPVCAGITGDLVSVINSPKLSVQKLGISY
SINPEASIPFGHPRVIGNEFDITTSKIPNTSGATGPGFASALDVCHEFIEIG
GRFVF"
BASE COUNT 430 a 153 c 205 g 475 t
ORIGIN
Query Match 24.6%; Score 206.6; DB 1; Length 1263;
Best Local Similarity 57.8%; Pred. No. 2.2e-27;
Matches 487; Conservative 0; Mismatches 339; Indels 17; Gaps 6;
Qy 1 ATGAATTATAGAAAATTTCTAGTAAGAGCGCGTTAATCTCATTAATGCTCAATCTTACCA 60
Db 397 ATGAATTGCAAGAAAATTTTATAACAAGTACACTAATATCATTTAGTGTCAATTTTACCT 456
Qy 61 TATCAGTCTTTGAGATCTGTAGGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120
Db 457 GGTGTGCTCTTTCTGATGTAATACAGAAAGACAGAGCCAGT---AGGTAGTGTGTTTAC 513
Qy 121 ATTAGTCAAGATCAATCCAAATGATATACACTTTTGAATAATCTCTGCTGGAAGAACT 180
Db 514 ATCAGTCAAAATACATGCCACCTGCTTCACTTTTGGTAAATGTCNAATTAAGAAGAT 573
Qy 181 CCTATTAAAT--GGAAACAAATTTCTCTCACTAAATAAGTTTTCGGACTAAAGAAGATGTTGA 239
Db 574 TCTAGAGATACATAAGTAGTATTGTTGCTTAAATAAGATTGGGATGAGTTAAACAAGT 633
Qy 240 TATAACAAAAGACGATTTTACAAGATAGTCCAGGCAATGATTTTCAATAAATCTT 299
Db 634 AGTAGCAATACAA---TTTTCACTGAAAAGATTATTTCTTTAAATATGAAAATAATATCC 689
Qy 300 AATATCAGGATTTTTCAGGAAGTATTGTTACTCTATGACGACCAAGAAATAGAACTTGA 359
Db 590 ATTCTTAGTGTGCTGGAGCAATTTGATCTCAATGAATGCTCAAGAAATAGAAATTTGA 749
Qy 360 AGCTGCATATCAACAATTTAATCCAAAACACCGATAACAATGATGATCTGATAATGTTGA 419
Db 750 AATATCTATGAACACTTTTGATGTAAATAACCCAGGTGGTAACTATAAAAATGATGCCA 809
Qy 420 ATACTATAACATTTTG---CAATATCTGCTAAAGATGCAATGGAAGATCAGCAATATGT 476
Db 810 TATGATTTGTCCTTAGACACAGCAACATCATCTGTTGGAGCAGCAGCAAGTACATCTGT 869
Qy 477 AGTACTTAAATAACGCGCATAACTTTTATGTCATTTGATGTTAATCTTCTCTGATGACAT 536

```



```
Db 870 TATGGTAAATAAATGAAAAATTTAAACAGATATATCATTAATGTTAAATGCATGTTATGATAT 929
Qy 537 TACAGCTGAAGGAGTATCTTTTCGTACCATATGCGATGTCAGGTATAGGAGCAGATCTTAT 596
Db 930 AATCGTGTAGTGAATGCCAGTGTCTCCATATGTAATGTCAGGTATAGGACCGATTTAGT 989
Qy 597 CACTATTTTAAAGACCTCAATCTAAATTTTCTTACCAAGGAAAAATAGGTATTTAGTTA 656
Db 990 GTCAGTAATTAATCTACAAATCTAAATTTATCTATCAAGGAAAACTTGGTATAAGTTA 1049
Qy 657 CCCTATCACACAGAGTCTCTGATTTATTTTGGTGGATACCTACCATGCGGTTATTTGTTAA 716
Db 1050 TTCGATAAATCCGAAGCATCTATATTTATTTGGCGGACATTTTCATAGAGTTATAGGTAA 1109
Qy 717 TAAATTTGAGAAGA---TACTCTGAATAAATCTCTGTAGTATTA--AATGATGCTCTCA 770
Db 1110 TGAGTTCAAGATATTACTACTCTTCAAGATTTTAACTAGTAACTGCTGGTGGCAC 1169
Qy 771 AACCATCTCTCTTCAGTAACCTCTGAGCTTTGAGTATCTTTTGGCGGAGAAATTTGGAATGAG 830
Db 1170 TCCAGGCTTTGCATCAGCAATCTTGTATGCTGCTGCAATTTTCGGTATAGAAATTTGGAGGAAG 1229
Qy 831 GTT 833
Db 1230 GTT 1232

RESULT 36
AX042313
LOCUS AX042313 837 bp DNA linear PAT 23-NOV-2000
DEFINITION Sequence 9 from Patent WO0065063.
ACCESSION AX042313
VERSION AX042313.1 GI:11340973
KEYWORDS
SOURCE
ORGANISM Ehrlichia chaffeensis.
REFERENCE Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
AUTHORS Anaplasmataceae; Ehrlichia.
1 (bases 1 to 837)
Barbet,A.F., Bowie,M.V., Ganta,R.R., Burrige,M.J., Mahan,S.M.,
Mcquire,T.C., Rurangirwa,F.R., Moreland,A.L., Simbi,B.H.,
Whitmire,W.W. and Alleman,A.R.
Nucleic acid vaccines against rickettsial diseases and methods of
use
JOURNAL Patent: WO 0065063-A 9 02-NOV-2000;
UNIVERSITY OF FLORIDA (US)
FEATURES
source Location/Qualifiers
BASE COUNT 279 a 130 c 159 g 269 t
ORIGIN
Query Match 24.4%; Score 205; DB 6; Length 837;
Best Local Similarity 55.3%; Pred. No. 4.8e-27;
Matches 466; Conservative 0; Mismatches 365; Indels 12; Gaps 3;
Qy 1 ATGAATATAGAAAATCTAGTAGAAGCGGTTATCTCAATTAATGTCATCTTACCA 60
Db 3 ATGAATTTGCAAAAATTTTTTATAACAACCTGCATTTAGTATCATCAATGTCCTTCTACCT 62
Qy 61 TATCAGTCTTTTGCAGATCTCTGTTAGTTCAGAACTATATGATAACAAGAGGCTTCTAC 120
Db 63 GGAATATCAATTTCTGATCCAGTGCA-----AGTGCAATATATAGTGGTAATTTCTAT 116
Qy 121 ATTAGTCAAAAGTACAATCCAAAGTATATCACACTTTTGAATAATTTCTGCTGGAAGAACT 180
Db 117 GTTAGTGGCAAGTATATGCCAAGTCTTCGCAATTTTGGCAATGTTTCTGCCAAGAAGAA 176
Qy 181 CCTATTATGAACAATAATCTCTCACTAAAAGTTTTCGGACTAAGAAGATGGTGAT 240
Db 177 AAAATCTCTACTGTTGCAATTTGATGGCTTTAAACAAGATTGGGAAGGATTAGCTCATCA 236
```

```
Qy 241 ATACAAAAAAGACGATTTTCAAGAGTAGCTCCAGGCATTTGATTTTCAAAATAACTTA 300
Db 237 AGTCACAATGATAATCATTTCAATAACAAGGGTTATTTCATTTAAATATGAATACCCCA 296
Qy 301 ATATCAGGATTTTTCAGGAAGTATTTGGTACTCTATGACCGACCAAGAAATAGAACTTGAA 360
Db 297 TTTTATGGGTTTGCAGGAGCTATTTGGTTATTTCATGGTGGTCCAGAGTAGAGTTTGA 356
Qy 361 GCTGCATATCAACAATTTAATCCAAAAACACCGATAACAATGATCTGATAATGGTGAA 420
Db 357 GTGCTCTATGAACATTTTGAGCTTTAAANAATCAGGGTAAATAACTATAAATATGATGCTCAC 416
Qy 421 TACTATAAACAATTTTGCATTTATCTCGTAAAGATGCAATG---GAAGATCAGCAATATGTA 477
Db 417 AGATACCTGCTGTTTGGTCAACAGACACAGCGGAATACCTAAACACTAGTAAATACGTA 476
Qy 478 GTACTTAAAAATGACGCATAAATTTTATGTCATTGATGGTTAATACTTGTCTATGACATTT 537
Db 477 CTGTTAAAAACCGAAGGATTTGCTTGACATATCATTTTATGCTAAATGCAATGCTATGATA 536
Qy 538 ACAGCTCAAGGAGTATCTTTTCGTACCATATGTCAGTGCAGGTATAGGAGCAGATCTTATC 597
Db 537 ATAAACGAGAGCATACCTTTGCTCTCTTACATATGTCAGGTGTTGGTACTGATTTAATA 596
Qy 598 ACTATTTTAAAGACCTCAATCTAAAAATTTGCTTTACCAAGGAAAAATAGGTATTAGTTAC 657
Db 597 TCCATGTTTGAAGCTACAAATCTTAAATTTCTTTACCAAGGAAGTTAGGTCTAAGTTAC 656
Qy 658 CCTATCACACAGAGTCTCTGCATTTTATGTCGGATGATCTACCATGGCGTTATTTGGTAAAT 717
Db 657 TCTATAAACCCAGAGCTCTCTGATTTTATTTGGTGGACATTTTTCATAAAGGTGATAGGAAC 716
Qy 718 AATTTTCAGAGATACCTGTAATAACTCTCTGATGATTTAATGATGTCCTCAACACCA 777
Db 717 GAATTTAGGACATTTCTCTGAAAGCATTTGTTAGTCATCAGCTACTCCAGATCTA 776
Qy 778 TCTGCTTTCAGTAACTCTTGACGTTGGATCTTTTGGCGGAGAAATTTGGAATGAGGTTCACC 837
Db 777 ---GCAATAGTAACTAAGTATGTCATTTTGGATAGAACTTTGGAGGAAGGTTTAAC 833
Qy 838 TTC 840
Db 834 TTC 836
```

RESULT 37  
AX042307

```
LOCUS AX042307 842 bp DNA linear PAT 23-NOV-2000
DEFINITION Sequence 3 from Patent WO0065063.
ACCESSION AX042307
VERSION AX042307.1 GI:11340967
KEYWORDS
SOURCE Ehrlichia chaffeensis.
ORGANISM Ehrlichia chaffeensis
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
1 (bases 1 to 842)
Barbet,A.F., Bowie,M.V., Ganta,R.R., Burrige,M.J., Mahan,S.M.,
Mcquire,T.C., Rurangirwa,F.R., Moreland,A.L., Simbi,B.H.,
Whitmire,W.W. and Alleman,A.R.
Nucleic acid vaccines against rickettsial diseases and methods of
use
JOURNAL Patent: WO 0065063-A 3 02-NOV-2000;
UNIVERSITY OF FLORIDA (US)
FEATURES
source Location/Qualifiers
1 .842
/organism="Ehrlichia chaffeensis"
/db xref="taxon:945"
1 .>840
/notes="unnamed protein product"
/codon_start=1
/transl_table=11
CDS
```

/protein\_id="CAC17073.1"  
/db\_xref="GI:11340968"  
/translation="MNYKKSFTAIIDINILLPGVSPDPRQVWVINGNYIYSKGYD  
AKASHFGVSPAKBERNTTVGFLKQNDWGSAINSPNDFTVSNYSFKYENPFLG  
PAGAIGYSDGPRIELEVYETFDVQNGNNYKNEAHRYCALSHNSAAMSSANSNFV  
FLKNEGLDIDIFMLNACVDVVEGIPRSPYICAGITDLVSMFEATNPKISYQKIGL  
SYSISPEASVFIGHFHFKVIGNEPRDIFTIIPGSLAGKNGYPAIVLVDVCHFGIEM  
GGRF"

BASE COUNT 282 a 143 c 166 g 251 t  
ORIGIN

Query Match 24.3%; Score 203.8; DB 6; Length 842;  
Best Local Similarity 57.8%; Pred. No. 7.8e-27;  
Matches 424; Conservative 0; Mismatches 297; Indels 12; Gaps 3;  
Qy 115 TTCTACATTAGTCGCAAGTACCAATCAAGTATATACACACTTTAGAAAAATTTCTGCT--- 171  
Db 109 TTCTACATCAGTGGAAATACAGTCCCAAGGCTTCGCATTTTGGAGTATTTCTGCTAAG 168  
Qy 172 GAAGAACTCTTATTAATGGAAACAAATTTCTCTCACTAAAAAGTTTTCGGACTAAAGAAA 231  
Db 169 GAAGAAAGAAATACAACTAGTTGGAGTGTTCGACTGAAGCAAAATTTGGACGGAGCGCA 228  
Qy 232 GATGGTGATATAACAAAAAGACCATTTTACAAGAGTAGTCCAGGCAATTCATTTTCAA 291  
Db 229 ATATCCAACTCTCCCAACAGATGATTTCACTCTCAATTTATTTAAATATGAA 288  
Qy 292 AATAACTTAAATATCAGGATTTTCAGGAAGTATTTGTTACTCTATGACGCGACCAAGATA 351  
Db 289 AACAAACCCGTTTTAGGTTTTGCAGGAGCTATTGTTACTCAATGATGTTGCCAAGATA 348  
Qy 352 GAACCTGAAGCTGCATATCAAAATTTTAAATCAAAAAACCCGATAACAAATGATCTGAT 411  
Db 349 GAGCTTGAAGTATCTTATGAACATTTTGATGTAAAAAATCAAGGTAAACAATTTAAGAA 408  
Qy 412 AATGGTGAATACTATAAACAATTTTGCATTTATCTGCTAAAGATGCAATG-----GAAGAT 465  
Db 409 GAAGCACAATAGATATGTGCTCTATCCATAACTCAGCAGCAGATAGTAGTCAAGT 468  
Qy 466 CAGCAATATGTAGTACTTTAAAAATGACGCGCAATACTTTTATGTCAATGATGTTAATACT 525  
Db 469 AATAATTTTGTCTTTCTAAAAAATGAAGGATTTACTTGACATATCATTTATGCTGAACGCA 528  
Qy 526 TGCTATGACATACAGCTGAGGAGTATCTTTCGTAACATATGCTATGCTGAGGATAGGA 585  
Db 529 TGCTATGACGTAGTAGGCGAAGGCATACCTTTTCTCTTATATATGCGCAGGTATCGGT 588  
Qy 586 GCAGATCTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGAAAAATA 645  
Db 589 ACTGATTTAGTATCCATGTTTGAAGCTACAATCTCAAAATTTCTTACNAGGAAGTTA 648  
Qy 646 GGTATTAGTTACCCATCACACAGCAAGTCTCTGCATTTATTTGGTGGATACCTACCATGCG 705  
Db 649 GGTTTAAGCTACTCTATAAGCCCAAGCTTCTGTGTTTATTTGGTGGCAGCTTTTCATAAG 708  
Qy 706 GTTATTGTAATAAATTTAGNAGATAC---CTGTAATAACTCTCTGTAGTATTAATGAT 762  
Db 709 GTAAATAGGGAACGAATTTAGAGATATTCCTACTATAAATACCTACTGATCAACACTTGA 768  
Qy 763 GCTCTCTCAAAACACATCTGTTCACTACTCTTGACGTTGGATATCTTTGGCGGAGAAAT 822  
Db 769 GGAAGGAAGAACTACCTCGCAATAGTATAGTGGATGATGCGCACTTTGGAAATAGAAATG 828  
Qy 823 GGAATGAGGTTCA 835  
Db 829 GGAGGAAGGTTTA 841

RESULT 38  
CRU49843  
LOCUS  
DEFINITION Cowdria ruminantium major antigenic surface protein (map1) gene, complete cds.  
linear BCT 01-MAY-1996

U49843  
U49843.1 GI:11292953  
heartwater rickettsia strain=Welgevonden stock.  
Ehrlichia ruminantium  
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
Anaplasmataceae; Ehrlichia.  
1 (bases 1 to 1564)  
Fehrsten, J., de Villiers, E.P., Brayton, K.A., van Kleef, M. and  
Allsopp, B.A.  
Direct Submission  
Submitted (23-FEB-1996) Etienne P. de Villiers, Molecular Biology,  
Onderstepoort Veterinary Institute, Onderstepoort 0110, South  
Africa  
Location/Qualifiers  
1. .1564  
/organism="Ehrlichia ruminantium"  
/strain="Welgevonden stock"  
/db\_xref="taxon:779"  
610. .1482  
/gene="map1"  
610. .1482  
/gene="map1"  
/note="MAP1; Cr32"  
/codon\_start=1  
/transl\_table=11  
/product="major antigenic surface protein"  
/protein\_id="AAA98610.1"  
/db\_xref="GI:11292954"  
/translation="MNCKKIPITSLVSLFPGVSPDVIQDSNPAGSVYISAKY  
MPTASHFGKMSIKEDSKNTOTVGLKKDWDGKVPKTSNTNYSLSLTKEDYSFRYEN  
PFLGPAAGIYSGMNGPRIELEVYETFDVQNGNNYKNEAHRYCALDHAQOASATNG  
LASSVMKNNENLTNLSMLNACVDVVEGIPRSPYICAGITDLVSMFEATNPKLSYQ  
KGKGISIINSEASIFIGHFHFKVIGNEPRDIFTIIPGSLAGKNGYPAIVLVDVCHFGI  
VCHFGIIBIGGRFIP"  
BASE COUNT 525 a 216 c 251 g 572 t  
ORIGIN  
Query Match 24.0%; Score 201.2; DB 1; Length 1564;  
Best Local Similarity 56.5%; Pred. No. 1.9e-26;  
Matches 493; Conservative 0; Mismatches 343; Indels 36; Gaps 5;  
Qy 1 ATGAATTAATAAGAAAAATTTCTAGTAAGAGCGCGTTAATCTCAATTAATCTCAATCTTACCA 60  
Db 610 ATGAATTCAGAAAAATTTTATCACAAGTACACTATATCATCTAGTCTATTTTACT 669  
Qy 61 TATCAGCTTTTTCAGATCTCTAGTCTTCAAGAACTAATGATAACAAAGAGGCTTCTAC 120  
Db 670 GGTGTGCTCTTTCTGA---TGTAATACAGGAGGATAGCAACCCAGCTGGTAGTGTTC 726  
Qy 121 ATTAGTCAAAAGTACAATCCAAAGTATATCACACTTTAGAAAAATTTCTCTGCTGAAGAACT 180  
Db 727 ATTAGCGCAAAATACATGCCAACTGCTTCACTTTGGTAAATGTCAATTAAGAAGAT 786  
Qy 181 CCTATTAT--GGAACTAATTTCTCTCACTAAAAAGTTTTCGGACTAAAGAAAGATGGTG 238  
Db 787 TCAAAAAATACTCAACAGTGTGTTGGTCTAAAAAAGATTTGGATGGAGTTAAAGTACCA 846  
Qy 239 ATATAACAAAAAAGA-----CGATTTTACAAGAGTAGCTCCAGGCAATGATTTT 288  
Db 847 ACATCAGAAAAACCAATTTACTCTTCACTTTTACTGAAAAAAGATTTATTTTCAGATAT 906  
Qy 289 CAAAATACTTAATATCAGGATTTTTCAGGAAGTATTTGTTACTCTATGCGCGGACCAAGA 348  
Db 907 GAAACCAATCCGTTTTCAGGTTTTCGTCGGCAATTTGGGTACTCAATGAATGGACCAAGA 966  
Qy 349 ATAGAATCTTGAAGCTGCATATCAAAATTTTAAATCCAAAAAACCCGATAACAAATGATCT 408  
Db 967 ATAGATTCGAAGTATCTCTATGAACCTTTTGTATGTAATAAAACCCAGGTGGCAACTACAA 1026  
Qy 409 GATAATGGTGAATACTATPAAACATTTTGCATTTATCTCTGTAAGATGCAATGGAAGATCAG 468  
Db 1027 AATGATGCACACATGCTACTGTGCTTTAGATACAGCACACAGAGTGCTACTTAATGGTGA 1086

QY 469 C-----AATATGTAGTACTTAAATAATGACGGCATAACTTTTATGTCTATTGATGGTT 519  
 Db 1087 ACATTAGCTTCATCTGTTATGATATAAATAATGAAAATTTAAACAATATATATCAATTAATGTTA 1146  
 QY 520 AATACCTTGCTATGACATTACAGCTGAAGGAGTATCTTTCGTACCATATGCAATGTCAGGT 579  
 Db 1147 AATGGCTTATGATATAATGCTGATGGAATGCCAGTTTCCATATGATGTCAGGT 1206  
 QY 580 ATAGAGCAGATCTTATCACTATTATTTAAAGACCTCAATCTAAATTTTGTCTTACCAAGGA 639  
 Db 1207 ATTGCTACTGATTAGTGTAGTAATTAATGCTACAAATCTAAATTTATCTTATCAAGGA 1266  
 QY 640 AAAATAGTATTAGTTACCTTATCACACCAAGAGTCTCTGATTTATTTGGTGGATAC 699  
 Db 1267 AAGCTAGGCATAAGTTACTCAATCAATCTGAAGCTTCTATCTTTATCGTGGACATTTTC 1326  
 QY 700 CATGCGCTTATTGGTAATAAATTTGAGAAGATACCTGTAATAA-----CTCCT 747  
 Db 1327 CATAGATTATAGTAATGAATTTAAGATATGCTACTTTAAATAATTTACCGCACT 1386  
 QY 748 GTAGTATTAAATGATGCTCTCAACACCACTCTGCTTCAGTAACCTTTGAGCTTGGATAC 807  
 Db 1387 AATRAAGTATCTACTGTAGCCATCCAGGTTTTCATCAGCAACACITGATGTTGTAC 1446  
 QY 808 TTGCGGAGAAATTTGGAATGAGTTTCACTT 839  
 Db 1447 TTTGGTATAGAAATTTGGAGGAAGGTTTATATT 1478

## RESULT 39

CRU50831  
 LOCUS 1263 bp DNA linear BCT 14-JUL-1996  
 DEFINITION Cowdria ruminantium Crystal Springs major antigenic protein 1  
 (map1) gene, complete cds.  
 ACCESSION U50831  
 VERSION U50831.1 GI:1418251  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 heartwater rickettsia strains-Crystal Springs.  
 Ehrlichia ruminantium  
 Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 Anaplasmataceae; Ehrlichia.  
 1 (bases 1 to 1263)  
 Reddy,G.R., Sulsona,C.R., Harrison,R.H., Mahan,S.M., Burridge,M.J.  
 and Barbet,A.F.  
 Cowdria ruminantium isolates from different geographical areas  
 Clin. Diagn. Lab. Immunol. 3 (4), 417-422 (1996)  
 96400830  
 8807206  
 2 (bases 1 to 1263)  
 Reddy,G.R.  
 Direct Submission  
 TITLE  
 Submitted (06-MAR-1996) Roman G. Reddy, Pathobiology, University of  
 Florida, 471 Mowry Rd., Gainesville, FL 32611-0880, USA  
 JOURNAL  
 MEDLINE  
 PUBLISHED  
 REFERENCE  
 FEATURES  
 source  
 1. 1263  
 /organism="Ehrlichia ruminantium"  
 /strain="Crystal Springs"  
 /db\_xref="taxon:779"  
 381..1244  
 /gene="map1"  
 381..1244  
 /gene="map1"  
 /note="surface protein"  
 /codon\_start=1  
 /translation="MNCCKIFITSLISLVFLPGVPSFSDVIQEDSNPAGSVISAKY  
 MPTTSHFGKMSIKDSKNQTVPLGKKWDGKVPKPSDSSNTNFTIFEKDYSRYENN  
 PPSGFAIGAIGYMNPGIGIEFVSYYETFDVKNPGNYKNNAHMYCALDTAAQNSTNGAG  
 1263  
 /product="major antigenic protein 1"  
 /protein\_id="AAC44142.1"  
 /db\_xref="GI:1418252"  
 /translation="MNCCKIFITSLISLVFLPGVPSFSDVIQEDSNPAGSVISAKY  
 MPTTSHFGKMSIKDSKNQTVPLGKKWDGKVPKPSDSSNTNFTIFEKDYSRYENN  
 PPSGFAIGAIGYMNPGIGIEFVSYYETFDVKNPGNYKNNAHMYCALDTAAQNSTNGAG  
 1263

LTTSVMVKNENLTNLSMLNACVDIMLDGIPVSPYVCAGIGTDLVSVINATNPKLKSYQ  
 GKUGISVINSINSEAIPIFGHFHRRVIGNEFKDIATIKLIFTSKTGISNPGFASATLDVCH  
 FGIEIGGRFVS"  
 BASE COUNT 440 a 178 c 193 g 452 t  
 ORIGIN

Query Match 23.7%; Score 199.4; DB 1; Length 1263;  
 Best Local Similarity 56.5%; Pred. No. 4.2e-26;  
 Matches 484; Conservative 0; Mismatches 346; Indels 27; Gaps 5;  
 QY 1 ATGAATTTATAAGAAAATTTCTAGTAAGAAGCGGTTAATCTCATTAATGTCTCAATCTTACCA 60  
 Db 381 ATGAATTTGCAAGAAAATTTTATCAAGTACACTATATCATTTAGTGTCTATTTTACCT 440  
 QY 61 TATCAGCTTTTTCAGATCCTGTAGGTTCAAGAACTAATGATACAAAGAGGCTTCTAC 120  
 Db 441 GGTGTGCTTTCTGATGTAATACAGAAAGACAGCAACCCAGC---AGGCAGTGTTTAC 497  
 QY 121 ATTAGTCAAAAGTACAACTCAAGTATATCACACTTTTAGAAAATTTCTCTGCTGAAGAACT 180  
 Db 498 ATTAGCGCAAAATACATGCCAACTACATCATTTTGGTAAATGTCAATCAAAAGAAAT 557  
 QY 181 CCTATTAAAT--GGAACAAATTTCTCTCACTAAAAAGTTTTCGG-----ACTAAAG 228  
 Db 558 TCAAAAATTAATCTCAACAGTATTTGGTCTAAAAAAGATTTGGGATGGCGTTAAACACCA 617  
 QY 229 AAAGATGGTGATATAACAAAAAAGAGCATTTTACAAAGAGTAGCTCCAGGCAATGATTTT 288  
 Db 618 TCAGATTTCTAGCAATCTAATTTCTACAAATTTTACTGAAAAGACTATTCTTTTCAGATAT 677  
 QY 289 CAAAATTAATTAATATCAGGATTTTTCAGGAAGTATTTGGTTACTCTATGAGCGGACCAAGA 348  
 Db 678 GAAAACAATCCGTTTTTCAGGTTTCGGTGGAGCAATTTGGGTACTCAATCAATGGAACCAAGA 737  
 QY 349 ATAGAATCTTGAAGCTGCATATCAACAATTTAATCCAAAAACCCGATACAAATGATACT 408  
 Db 738 ATAGATTTGGAAGTATCTTATGAACTTTTGTATGTAAAAAACCCAGGTGGCAACTATAA 797  
 QY 409 GATAATGGTGAATACTATAAACATTT-----TGCATTATCTCTGTAAGATGCAATG 459  
 Db 798 AACAAACGACACATGATGCTGCTTTAGATACAGCAGCACAAAATAGCATAATGCGCCA 857  
 QY 460 GAAGATCAGCAATATGATGATGCTTTAAAAATGAGCGGCAATACTTTTATGTCATGATGTT 519  
 Db 858 GGATTAATCACTACTGTTATGTTAAAAAAGCAAAAATTTAAACAATATATCATTAATGTTA 917  
 QY 520 AATACTTGTCTATGATACATGATGAGGAGTATCTTTCGTACCATATGATGTCAGGT 579  
 Db 918 AATGCGTGTATGATATCATGCTTGTATGTAATACCAAGTTCTTCCATATGATGTCAGGT 977  
 QY 580 ATAGGAGCAGATCTTATCACTATTTTAAAAAGCCTCAATCTAAAAATTTGCTTACCAAGGA 639  
 Db 978 ATTGGCACTGACTTAGTGTGCTAGTAATTAATGCTCAACAATCTTAATTTCTTATCAAGGA 1037  
 QY 640 AAAATAGGTAATGATTACCTTATCACACCAAGAGTCTCTGCAATTTATTTGGTGGATAC 699  
 Db 1038 AAGCTAGGCATAAGTTACTCAATCAATCTGAAAGTCTTCTTTATCGTGGACATTTTC 1097  
 QY 700 CATGGCGTTATTGGTAATAAATTTGAGAAGATACCT---GTAATACTCTCTGATGATTA 756  
 Db 1098 CATAGAGTTATAGGTAATGAATTTAAAGATATTGCTACCTTAAAAATTTTACTTCAAAA 1157  
 QY 757 AATGATGCTCTCTCAAAACCACTATGCTTTCAGTAACTCTTTCAGCTTGGATATCTTTGGCGGA 816  
 Db 1158 ACAGGAATATCTAATCTGCTGGCTTTGCATCAGCAACTGATGTTTGTCTCATTGGTATA 1217  
 QY 817 GAAATTTGGAATGAGGTT 833  
 Db 1218 GAAATTTGGAAGGTT 1234

RESULT 40  
 AY028378

LOCUS AY028378 873 bp DNA linear BCT 06-NOV-2001  
 DEFINITION Cowdria ruminantium major antigenic protein MAP1 gene, complete  
 cds.  
 ACCESSION AY028378  
 VERSION AY028378.1 GI:13378319  
 KEYWORDS  
 SOURCE Ehrlichia ruminantium.  
 ORGANISM Ehrlichia ruminantium  
 Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 Anaplasmataceae; Ehrlichia.  
 REFERENCE 1 (bases 1 to 873)  
 AUTHORS Allsopp, M.T., Dorfling, C.M., Maillard, J.-C., Bensaïd, A.,  
 Haydon, D.T., van Heerden, H. and Allsopp, B.A.  
 TITLE Ehrlichia ruminantium major antigenic protein gene (map1) variants  
 are not geographically constrained and show no evidence of having  
 evolved under positive selection pressure  
 J. Clin. Microbiol. 39 (11), 4200-4203 (2001)  
 JOURNAL 21539003  
 MEDLINE 11682561  
 PUBMED  
 REFERENCE 2 (bases 1 to 873)  
 AUTHORS Bensaïd, A., Allsopp, M.T., Maillard, J.-C., Chantal, I. and Allsopp, B.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-MAR-2001) Elevage et Medecine Veterinaire, Centre de  
 Cooperation Internationale en Recherche Agronomique pour le  
 Developpement., Domaine Duclos, Petit Bourg 97170, Guadeloupe  
 FEATURES  
 source  
 1..873  
 /organism="Ehrlichia ruminantium"  
 /strain="Vosloo"  
 /db\_xref="taxon:779"  
 /note="obtained from cultured organisms"  
 CDS  
 1..873  
 /note="outer membrane protein"  
 /codon\_start=1  
 /transl\_table=1  
 /product="major antigenic protein MAP1"  
 /protein\_id="AAK18727.1"  
 /db\_xref="GI:13378320"  
 /translation="MNCKKIFITSLISLFLPQVSDVIQSDNPAGSVYISAKY  
 MPTASHFGKMSIKEDSKSTQVFLKMDGKVPKPTSENTNYSLSFTEKDYSPRYENN  
 PFLGPAIGYVMNGPRILEFVSVEYVDKPNPGNYKNDAMHICALDTAQQSANTGAT  
 LASSVMIKENLNIISLMNLAYDMLDMPVSPVACAGIGTDLVSNATNPKLISYQ  
 GKLGISYINSEASIFIGGHRFVIGNEFKDIATIKIFTATNKYSTVANPFPASATLD  
 VCHGIEIGGRFV"  
 BASE COUNT 303 a 134 c 155 g 281 t  
 ORIGIN  
 Query Match 23.4%; Score 196.8; DB 1; Length 873;  
 Best Local Similarity 56.4%; Pred. No. 1.4e-25;  
 Matches 488; Conservative 0; Mismatches 342; Indels 36; Gaps 5;  
 QY 1 ATGAATTATAGAAATTTCTAGTAGAGAGCGGTTAACTCTCAATTAATGTCATCTTACCA 60  
 DB 1 ATGAATTCGAAGAAATTTTATATCAAGTACACTAATATCATTTAGTGTCTTTTACCT 60  
 QY 61 TATCAGTCTTTTGCAGATCCTGTAGTTCACAGAACTAATGATATACAAAGAGGCTTCTAC 120  
 DB 61 GGTGTCTCTTTTCTGA---TGTAATACAGGAGGATAGCAACCCAGCTGGTAGTGTTTAC 117  
 QY 121 ATTAGTGCAGAAAGTACAATCCAGATATACACTTTTAGAAATTTCTCTGCTGAAGAACT 180  
 DB 118 ATTAGCGCAAAATACATGCCAACTGCTTCACACTTTTGGTAAATATGTCATTAAGAAGAT 177  
 QY 181 CCTATTA--ATGGAAACAAATTTCTCTACATAAAAAAGTTTTCGGACTAAAGAAAGATGGTG 238  
 DB 178 TCAAAAAGTACTCAAAACAGTGTGTGCTCTAAAAAAGATTTGGATGGAGTTAAAGTACCA 237  
 QY 239 ATATACAAAAAAGA-----CGATTTTACAAGAGTAGTCCAGGCAATGATTTT 288  
 DB 238 ACATCAGAAACACACCAATTTACTCTTTCTTTTACTGAAAAGATTTATTTCTTCAGATAT 297  
 QY 289 CAAATAACTTAATACAGGATTTTTCAGAGATTTTGGTTACTCTATGAGCGACCAAGA 348

Db 298 GAAACCAATCGTTTTTTAGGTTTTTCTGGAGCAATTTGGGTACTCAATGAATGGACCAAGA 357  
 QY 349 ATAGAACTTGAAGCTGCATATCAACAATTTTAATCCAAAAAACACCGATAAACAATGATACT 408  
 DB 358 ATAGAGTTTGAAGTATCTCTATGAAACTGTTGATGTAATAAAACCCAGGTGGCAACTACAAA 417  
 QY 409 GATAATGGTGAATACTATATAAATTTTGCATTTATCTCGTAAAGATGCAATGGAAGATCAG 468  
 DB 418 AATGATGCACACATGTTACTGTGCTTTAGATACACACAACAGAGTGTCTATTAATGGTGCA 477  
 QY 469 CA-----ATATGTAGTACTTTAAAAATGACCGGCATACTTTTATGTCAATTTGATGGTT 519  
 DB 478 ACATTAGCTTCATCTGTTATGATATAAAATGAAATTTAACAAATATATCATTAATGTTA 537  
 QY 520 AATACTTGTATGATACATTACAGCTGAAGGAGTATCTTTTCGTAACCATATGCAATGTCAGGT 579  
 DB 538 AATCGTGTATGATATAAATGCTTGTATGGAATGCGAGTTTCTCCATATGATGTGCGAGGT 597  
 QY 580 ATAGAGCAGATCTTATCAGTATTTTAAAGACCTCAATCTAAATTTGCTTACCAGGA 639  
 DB 598 ATTGGTACTGATTTAGTGTGCTAGTAAATTAATGCTACAAATCTTAATTTATCAAGGA 657  
 QY 640 AAAATAGGTATTTAGTTACCTTATCACACAGAGTCTCTGCAATTTATTTGGTGGATACTAC 699  
 DB 658 AAGCTAGGCATAAGTCTTCAATCAATTTCTGAAGCTTCTATCTTTATCGGTGGACATTTT 717  
 QY 700 CATGGCGTTATGTTAATAAATTTGAGAAGATACCTGTATAATA-----CTCCT 747  
 DB 718 CATAGAGTTATAGGTAATGAATTTAAAGATATTGCTACTTTTAAAAAATATTTTACCGCACT 777  
 QY 748 GTAGTATTAATGATGCTCTCAACCCACATCTGCTTCAGTAACTCTTGAGTTGGATAC 807  
 DB 778 AATAAGTATCTACTGTAGCCAAATCCAGGTTTTCATCAGCAACACACTTGTGTTGTCAC 837  
 QY 808 TTTGCGCGAGAAATTTGGAATGAGGTT 833  
 DB 838 TTTGCTATAGAAATTTGGAGGAGGTT 863

Search completed: July 4, 2003, 00:43:40  
 Job time : 1664 secs

352 GAACCTTGAAGTCGATATCAACAATTAATCCAAAAACACGATACCAATGATCTAT 411  
 361 GAATTTGAAGTCGATACCAACAATTTGATGCAAAAAATCTGACAAATGACACTAT 420  
 412 AATGCGAATATCAATAAATTTGATATCTCGTAAGATGCAATGCAATGACAA 471  
 421 AGCGGTGCTACTATATAATATCTTGGACTATCTCGTGAAGACCAATACGATAGAAA 480  
 472 TATGTAGTATCTTAAAAATGACGCAATCTTTTATGTCATGATGTTAATCTGCTAT 531  
 481 TATGTTGCTCTTAAATAGAGCATCACTTTTATGTCATGATGTTAATGTTAATGCTAT 540  
 532 GACATTACAGTCGAGGAGTATCTTTTCGATACCATATGCTATGTCAGGATAGAGCAGAT 591  
 541 GACATTACAGTCGAGGAGTATCTTTTCATACCGTATGATGTCAGGATAGAGCAGAT 600  
 592 CTTATCACTATTTTAAAGACCTCAATCTTAAATTTGCTTACCAAGGAAATAGTAT 651  
 601 CTTATAAAGCTATTTAAGGATTTTAAATTTTAAATTTCTCATCCAGGAAATAGTAT 660  
 652 AGTATCCCTATCACACGAGCTCTGCAATTTTATGTTGGATCTACCATGCGGTTAT 711  
 661 AGTATCCCTATCACACGAGCTCTGCAATTTTATGTTGGATCTACCATGCGGTTAT 720  
 712 GGTAATAAATTTGAGAGGATACCTGTAAATACCTCTGTAGTATTAATGATGCTCTCAA 771  
 721 GGAATAATTTTAAACAAATACCTGTAAATACCACTGTAGTATTAAGAGGAGCTCTCAA 780  
 772 ACCATCTGCTTCAGTAACTCTTGAAGTGTGATCTTTGGCGGAGAAATGGAATGAG 831  
 781 ACCATCTGCTTCAGTAACTCTTGAAGTGTGATCTTTGGCGGAGAAATGGAATGAG 840  
 832 TTCACCTTC 840  
 841 TTCACCTTC 849

RESULT 3  
 AAX34770  
 ID AAX34770 standard; DNA; 495 BP.  
 AC AAX34770;  
 DT 05-JUL-1999 (first entry)  
 DE DNA encoding P30-10protein.  
 KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 KW detection; dog; ss.  
 OS Ehrlichia canis.  
 PN WO9913720-AL.  
 XX 25-MAR-1999.  
 XX 18-SEP-1998; 98WO-US19600.  
 XX 19-SEP-1997; 97US-0059353.  
 XX (OHIS ) UNIV OHIO STATE.  
 XX Ohashi N, Rikihisa Y;  
 XX WPI; 1999-254290/21.  
 XX P-PSDB; AAY06970.  
 XX Novel outer membrane proteins from Ehrlichia chaffeensis and  
 XX Ehrlichia canis  
 XX Disclosure; Fig 30A; 55pp; English.

CC The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1(B to 2) shown  
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and  
 CC consist of proteins shown in AAY06959-970. The proteins and genes are  
 CC used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX Sequence 495 BP; 192 A; 81 C; 80 G; 142 T; 0 other;  
 SQ

Query Match 53.7%; Score 450.8; DB 20; Length 495;  
 Best Local Similarity 96.8%; Pred. No. 6.1e-97;  
 Matches 482; Conservative 0; Mismatches 12; Indels 4; Gaps 2;

QY 1 ATGAATTAATAGAAAATTTCTAGTAAGAAGCGGTTAATCTCATTAATGTCATCTTACCA 60  
 DB 1 ATGAATTAATAGAAAATTTCTAGTAAGAAGCGGTTAATCTCATTAATGTCATCTTACCA 60  
 QY 61 TATCAGTCTTTTGGCAGATCCTGTAGTTCAAGACTAATGATAACAAAGAGGCTTCTAC 120  
 DB 61 TATCAGTCTTTTGGCAGATCCTGTAGTTCAAGACTAATGATAACAAAGAGGCTTCTAC 120  
 QY 121 ATTAGTCAAAAGTACAAATCTCAAGTATATCACACTTTAGAAAATTTCTGCTGAAGAACT 180  
 DB 121 ATTAGTCAAAAGTACAAATCTCAAGTATATCACACTTTAGAAAATTTCTGCTGAAGAACT 180  
 QY 181 CCTATTATGGAACAAATTTCTCTCACTAAAGTTTTCGGACTAAAGAAAGATGGTGAT 240  
 DB 181 CCTATTATGGAACAAATTTCTCTCACTAAAGTTTTCGGACTAAAGAAAGATGGTGAT 240  
 QY 241 ATAACAAAAAGAGCGATTTTACAGAGTAGCTCCAGGCAATGATTTTCAAAATTAACCTTA 300  
 DB 241 ATAACAAAAAGAGCGATTTTACAGAGTAGCTCCAGGCAATGATTTTCAAAATTAACCTTA 300  
 QY 301 ATATCAGGATTTTCAGGAAGTATTTGTTACTCTATGACGGACCAAGATAGAACTTGAA 360  
 DB 301 ATATCAGGATTTTCAGGAAGTATTTGTTACTCTATGACGGACCAAGATAGAACTTGAA 360  
 QY 361 GCTGCATATCAACAATTTAATCCAAAAACACCGATACCAATGATGATTAATGGTGAA 420  
 DB 361 GCTGCATATCAACAATTTAATCCAAAAACACCGATACCAATGATGATTAATGGTGAA 417  
 QY 421 TACTATAAACATTTTGGCATTATCTCTAAAGATGCAATGGAAGATCAGCAATATGTAGTA 480  
 DB 418 TACTATAAACATTTTGGCA-TAATCTGTAAGATGCCNTGGAAGATCAGCAATATGTGTT 476  
 QY 481 CTTAAAAATGACGGCATA 498  
 DB 477 CTTAAAAATGACGGCATA 494

RESULT 4  
 AAX34761  
 ID AAX34761 standard; DNA; 924 BP.  
 AC AAX34761;  
 DT 05-JUL-1999 (first entry)  
 DE DNA encoding P30-1 protein.  
 KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 KW detection; dog; ss.  
 OS Ehrlichia canis.  
 PN WO9913720-AL.  
 XX 25-MAR-1999.  
 XX 18-SEP-1998; 98WO-US19600.  
 XX 19-SEP-1997; 97US-0059353.

(OHIS ) UNIV OHIO STATE.  
 Ohashi N, Rikihisa Y;  
 WPI; 1999-254290/21.  
 P-PSDB; AAY06961.  
 Novel outer membrane proteins from Ehrlichia chaffeensis and Ehrlichia canis  
 Disclosure; Fig 21A; 55pp; English.  
 The invention provides isolated outer membrane proteins (OMP) from Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part of the OMP family and consist of proteins OMP-1, -1(B to Z) shown in AAY06943-958. The E. canis proteins form part of the p30 family and consist of proteins shown in AAY06959-970. The proteins and genes are used to detect E. chaffeensis in patients and E. canis in dogs.  
 SQ Sequence 924 BP; 301 A; 158 C; 160 G; 305 T; 0 other;  
 Query Match 28.4%; Score 238.4; DB 20; Length 924;  
 Best Local Similarity 57.7%; Pred. No. 8.8e-47;  
 Matches 486; Conservative 0; Mismatches 346; Indels 10; Gaps 3;  
 QY 1 ATGAAATTATAGAAAATCTAGTAAGAGCGCGTTATCTCAATTAATGCTCAATCTTACCA 60  
 DB 88 ATGAAATTGCAAAAATCTTATTAACAACACTGCATTAATATCAATTAATGCTCAATCTTACCA 147  
 QY 61 TATCAGTCTTTTTCAGATCTGTAGGTTCAAGACTAATGATACAAAGAGGCTTCTAC 120  
 DB 148 AGCATATCTTTTCTGATACTATACAGATGGTAACATGGTGGTAA-----CTTCTAT 201  
 QY 121 ATTAGTCAAAAGTACAAATCAAGTATATCAACATTTTAGAAAATCTCTGCTGAAGA--AA 178  
 DB 202 ATTAGTGAAGATGATGACCAAGTGTCTCACTTTTGGTAGCTTCTAGCTAAAGAGAA 261  
 QY 179 CTCCTATTATGGNACAAATCTCTCACTAAAAGTTTTCGGACTAAAGAAAGATGGTG 238  
 DB 262 AGCAAAATCAACTGTGTGGATTTTGGATTAAACATGATTTGGAGTGAAGTCCAACTACT- 320  
 QY 239 ATATAACAAAAGAGCGATTTTACAGAGTAGTCCAGGCAATGATTTTCRAATTAAT 298  
 DB 321 -TAAGATAAACACGCTGACTTTACTGTTCCAAACTATTGTTCCAGATACGAGAACATC 379  
 QY 299 TAATATCAGGATTTTCAGGAAGTATTGTTACTCTATGGACGACCAAGATAGAACTTG 358  
 DB 380 CATTCTAGGGTTTGCAGGAGCTATCGTTACTCAATGGTGGCCCAAGATAGATTG 439  
 QY 359 AGCTGATATCAACAATTTAATCCAAAACACCGATTAACATGATGATGATGATGATG 418  
 DB 440 AATATCTTATGAAGCATTCGACGTAAGAAATGCTTAATATCAATTAATCAAAATGACGCGC 499  
 QY 419 AATATATTAACATTTTGGCATTATCTCGTAAGATGCAATGGAGATCAGCAATATCTAG 478  
 DB 500 ACAGGTACTCGCTTATCTATCTATCATCACAATCGGCGACGATGGAAGCTGATAATTTGCT 559  
 QY 479 TACTTAAAATGACGGCATAACTTTTATGTCAATGATGTTAATATCTTGTATGACATTA 538  
 DB 560 TCTTAAAACGAAGGTTAATGACATATCACTTGGCAATAATGATGATGATGATATAA 619  
 QY 539 CAGCTGAAGGATATCTTTCGTACCATATGATGATGATGATGATGATGATGATGATGAT 598  
 QY 620 TAAATGACAAAGTACCTGTTTCTTCTTATATATGCGCAGGTATTGGTACTGATTTGATTT 679  
 QY 599 CTATTTTAAAGACCTCAATCTAAAATTTCTTACCAAGGAAATAGGATGATGATGATGAT 658  
 QY 680 CTATGTTGAGCTACAGTCTTAAATTTCTTACCAAGGAAATAGGATGATGATGATGAT 739  
 QY 659 CTATCACACCAAGTCTCTGCAATTTATTTGGTGATGATGATGATGATGATGATGATGAT 718  
 QY 740 CTATTAATCGGAAACCTCTGTTTTCATCGTGGGCAATTTCCACAGGATCATAGGTAATG 799

QY 719 AATTGAGAGATACCTGTATTAATACTCTCTGTAGTATTAAATGATGCTCTCAACCAAT 778  
 DB 800 AGTTTAGAGATATTCTCTGCAATAGTACCTAGTACTCACTCACTAATAAGTGGACCAAT 859  
 QY 779 CTGCTTCAGTAACCTCTTGGAGTTGGATACCTTGGCGGAGAAATTTGGAATGAGTTCACT 838  
 DB 860 TTGCACAGTAACACTAAATGTGTGTCTCACTTTGTTTGTAGAACTTTGGAGGAAGATTTAACT 919  
 QY 839 TC 840  
 DB 920 TC 921

RESULT 5  
 AAD01292  
 ID AAD01292 standard; DNA; 1607 BP.  
 AC AAD01292;  
 XX 12-OCT-2000 (first entry)  
 DT DT  
 XX Ehrlichia canis immunoreactive protein ECa28-1 DNA.  
 DE DE  
 XX Homologous 28-kDa protein gene; ECa28-1; immunoreactive; vaccine;  
 KW p28 gene; polymorphic multiple gene family; immunoprotective antigen;  
 KW antibacterial; canine ehrlichiosis; canine tropical pancytopenia;  
 KW tick-borne rickettsial disease; serodiagnosis; ds.  
 XX Ehrlichia canis.  
 OS  
 XX Key Location/Qualifiers  
 FH CDS 146..982  
 FT /tag= a  
 FT /product= "ECa28-1 protein (30-kDa)"  
 FT sig\_peptide 146..214  
 FT /tag= b  
 FT mat\_peptide 215..979  
 FT /tag= c  
 FT /product= "Mature ECa28-1 protein (28-kDa)"  
 XX  
 PN WO200032745-A2.  
 XX  
 PD 08-JUN-2000.  
 XX  
 PF 24-NOV-1999; 99WO-US28075.  
 XX  
 PR 30-NOV-1998; 98US-0201458.  
 PR 03-MAR-1999; 99US-0261358.  
 XX  
 XX (RERE-) RES DEV FOUND.  
 PA Walker DH, Yu X, McBride JW;  
 PI WPI; 2000-412258/35.  
 XX P-PSDB; AAY7147.  
 DR  
 DR Ehrlichia canis antigens useful for vaccinating against canine  
 PT ehrlichiosis in dogs -  
 XX  
 PS Claim 5; Fig 1; 86pp; English.  
 XX  
 CC The patent relates to homologous 28-kilodalton (kDa) protein genes of  
 CC Ehrlichia canis, designated ECa28A1, ECa28A2, ECa28A3, ECa28-1 and  
 CC ECa28-2. These genes are members of a polymorphic multiple gene family  
 CC and contained in a single locus of 5.592 kb. The 28-kDa proteins are  
 CC immunoreactive with anti-E. canis serum hence are important  
 CC immunoprotective antigens. The protein is useful for vaccinating  
 CC against E. canis infections such as canine ehrlichiosis in dogs.  
 CC Canine ehrlichiosis, also known as canine tropical pancytopenia, is a  
 CC tick-borne rickettsial disease of dogs. ECa28-1 is conserved amongst  
 CC different strains of E. canis and hence useful for serodiagnosis of  
 CC canine ehrlichiosis. The present sequence is a DNA encoding E. canis  
 CC ECa28-1 30-kDa protein which is post-translationally modified to a

ure 28-kDa protein by cleavage of N-terminal signal sequence.

Sequence 1607 BP; 543 A; 268 C; 250 G; 540 T; 6 other;

Query Match 28.4%; Score 238.4; DB 21; Length 1607;

Best Local Similarity 57.7%; Pred. No. 9.9e-47;

Matches 486; Conservative 0; Mismatches 346; Indels 10; Gaps 3;

```

QY 1 ATGAATTATAGAAAATCTAGTAAGAGCGGTTAATCTCATTAATGCTCAATCTTACCA 60
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
146 ATGAATTGCAAAAATCTTATACAACTGCATTAATATATTAATGCTCAATCTTACCA 205
QY 61 TATCAGCTTTTTCAGATCTGTAGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
206 AGCATATCTTTTCTGATCTATACAGATGTTCAAGATGGTGGTAA-----CTTCTAT 259
QY 121 ATTAGTCAAGTACAACTCAAGTATATACACTTTAGAAAATCTCTGCTGAAGA--AA 178
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
260 ATTAGTGGAAAGTATGACCAAGTGTCTCACATTTTGGTAGCTTCTCAGCTTAAAGAGAA 319
QY 179 CTCCTATTATGGAACAAATCTCTCACTAAAAAGTTTTCGGACTAAAGAAAGATGGT 238
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
320 AGCAATCACTGTGGATTTTGGATTAACATGATGGATGGAGTCCAACTACT- 378
QY 239 ATATACAAAAAGACGATTTTACAGAGTAGTCCAGGCAATGATTTTCAAAATAACT 298
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
379 -TAAGAATAAACAGCTGACTTTACTGTTCCAAATCTTCTGATACAGATACGAGAACATC 437
QY 299 TAATATCAGGATTTTCAGGAAGTATGTTACTCTNTGGAGCGACCAAGATAGAACTTG 358
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
438 CATTTCAGGGTTTTCAGGAGCTATCGGTACTCAATGGGTGGCCCAAGATAGAAATCG 497
QY 359 AAGCTGCATATCAACAATTTAATCAAAAACCGGATAACAATGATGATGATGATGATG 418
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
498 AATATCTTATGAAGCAATCGACGTAAAGAGTCTTAATATCAATATCAAAATGACGCGC 557
QY 419 AATACTATAACAATTTTGCATTTCTCGTAAGATGCAATGGAAGATCAGCAATATGATG 478
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
558 ACAGGTACTGGCTCTATCTCATCACACATCGGCAGCCATGGAGCTGATATAATTTGCT 617
QY 479 TACTTAAATAGCGGCATTAATCTTTTATGTCATTTGATGGTTAATACTTGTATGACATTA 538
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
618 TCTTAAAAAAGCAAGGGTAAATGACATATACACTTGCATTAATAATGCAATGATATATA 677
QY 539 CAGCTGAAGAGTATCTTTCGTACCATATGTCAGGTATAGGAGCAGATCTTATCA 598
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
678 TAATGACAAAGTACCTGTTCTCTTATATATGCGCAGGTATTTGGTACTGATTTGATTT 737
QY 599 CTATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAAGGAAAAATAGGTATTTAGTTACC 658
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
738 CTATGTTTGAAGCTACAACTCTAAATTTCTTACCAAGGAAAACTGGGCATTTAGTTACT 797
QY 659 CTATCACACAGAGTCTCTGCATTTATGTTGATGATACCATGCGGCTTATGTTAATA 718
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
798 CTATTAATCCGGAACCTCTGTTTTCATCGGTGGGCATTTCCACAGGATCATAGTAATG 857
QY 719 AATTGAGAGATACCTGTTAATACTCTGTAGTATTAAATGATGCTCCTCAACACCAT 778
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
858 AGTTTAGAGATATTTCTTGCATAGTACTAGTAACTCACTAATAGTGGACCAAT 917
QY 779 CTGCTTCAGTAATCTTTGACGTTGGATACCTTTGGCGGAGAAATTTGGAATGAGGTTACCT 838
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
918 TTGCAACAGTAACTAAATGTTGTGTCTACCTTTGTTGTTTGTAGAACTTTGGAGGAAGATTTAACT 977
QY 839 TC 840
DB ||
978 TC 979

```

RESULT 6

ABK68852

ID ABK68852 standard; DNA; 1607 BP.

XX

ABK68852;

02-JUL-2002 (first entry)

DNA encoding Ehrlichia canis p28-7.

Ehrlichia canis infection; vaccine; serodiagnostic; gene; p28; ss;  
antibacterial.

Ehrlichia canis.

WO200222782-A2.

21-MAR-2002.

12-SEP-2001; 2001WO-US28759.

12-SEP-2000; 2000US-0660587.

(RERE-) RES DEV FOUND.

Walker DH, Yu X, McBride JW;

WPI; 2002-351882/38.

P-PSDB; AAU96100.

New recombinant homologous 28 kilodalton immunodominant protein from  
Ehrlichia canis, useful for treating Ehrlichia canis infections

Claim 5; Figure 1; 106pp; English.

The invention relates to a recombinant homologous 28 kDa immunodominant  
protein, p28 (I) of Ehrlichia canis (I), a 28-kDa antigen preferably  
dispersed in a pharmaceutically acceptable carrier, is useful for  
inhibiting E. canis infection in a subject. (I) is useful in the  
development of vaccines and serodiagnoses that are particularly  
effective for disease prevention and serodiagnosis. ABK68852-ABK68878  
represent the 28-kDa antigen coding sequences and PCR primers of the  
invention.

Sequence 1607 BP; 543 A; 268 C; 250 G; 540 T; 6 other;

Query Match 28.4%; Score 238.4; DB 24; Length 1607;

Best Local Similarity 57.7%; Pred. No. 9.9e-47;

Matches 486; Conservative 0; Mismatches 346; Indels 10; Gaps 3;

```

QY 1 ATGAATTATAGAAAATCTAGTAAGAGCGGTTAATCTCATTAATGCTCAATCTTACCA 60
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
146 ATGAATTGCAAAAATCTTATACAACTGCATTAATATATCAATTAATGCTCAATCTTACCA 205
QY 61 TATCAGCTTTTTCAGATCTGTAGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
206 AGCATATCTTTTCTGATCTATACAGATGGTAACTGGTGGTAA-----CTTCTAT 259
QY 121 ATTAGTCAAAAGTACAACTCAAGTATATACACTTTTAGAAAAATTTCTGCTGAAGA--AA 178
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
260 ATTAGTGGAAAGTATGTACCAAGTGTCTACATTTTGGTAGCTTCTCAGCTTAAAGAGAA 319
QY 179 CTCCTATTATGGAACAAATCTCTCACTAAAAAGTTTTCGGACTAAAGAAAGATGGTG 238
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
320 AGCAATCACTGTGGAGTGTGGATTTAAACATGATTTGGGATGGAGTCCAACTACT- 378
QY 239 ATATACAAAAAAAGACGATTTTACAGAGTAGTCCAGGCAATTTTTCAAAATAACT 298
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
379 -TAAGAATAAACACGCTGACTTTTCTGTTCCAACTATTCTGTTCCAGATACGAGAACATC 437
QY 299 TAATATCAGGATTTTTCAGGAAGTATTTGGTTACTCTATGGACGGACCAAGATAGAACTTG 358
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
438 CATTTCAGGGTTTTCAGGAGCTATCGGTACTCAATGGGTGGCCCAAGATAGAAATCG 497
QY 359 AAGCTGCATATCAACAATTTAATCAAAAACCGGATAACAATGATGATGATGATGATG 418
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
498 AATATCTTATGAAGCAATTCGAGCTTAAAGAGTCTTAATATCAATATCAAAATGACGCGC 557

```

```

QY 419 AATACATATAACATTTTGGCATATCTCGTAAGATGCAATGAGATCAGCATATGTAG 478
Db 558 ACAGGTACTCGCGTCTATCTCATCACACATCGGCAGCCATGGAAGCTGATAAATTTGTCT 617
QY 479 TACTTAAAAATGACGGCATAACCTTTTATGTGTCATTGATGTTAATACTTGTATGACATTA 538
Db 618 TCTTAAAAAAGAGAGGTTAATGACATATCACTTGCATTAATGATGATGATATATAA 677
QY 539 CAGCTGAAGAGGATCTTTTCGTAACCATATGCGATGTCAGGTATAGGAGCAGATCTTATCA 598
Db 678 TAAATGACAAAGTACCTGTTTCTCTTATATATGCGCAGGTATTTGTTACTGATTTGATTT 737
QY 599 CTATTTTAAAGACCTCAATCTAAATTTTCTTACCAAGGAAAAATAGTATAGTTATACC 658
Db 738 CTATGTTGAAGCTACAAGTCTCTAAATTTTCTTACCAAGGAAAAATAGTATAGTTACT 797
QY 659 CTATCACACCAAGTCTCTGCAATTTATTTGGTGATCTACCATGCGGTTATTTGTAATA 718
Db 798 CTATTAATCCGGAACCTCTGTTTTCATCGGTGGGCAATTTCCACAGATCATAGGTAATG 857
QY 719 AATTTGAGAGATACCTGTAATAACTCTCTGTAGTATTAATGATGCTCTCAAAACACAT 778
Db 858 AGTTTAGAGATATTCCTGCAATAGTACCTAGTAACCTCAACTACAATAAGTGGACCAAT 917
QY 779 CTGCTTCAGTAATCTTGACGTTGGATCTTTGGCGGAGAAATTTGGAATGAGTTCACT 838
Db 918 TTGCAACAGTAACACTAAATGTGTGTCTCTTTTGTAGAACTTTGGAAGATTTAACT 977
QY 839 TC 840
Db 978 TC 979

```

## RESULT 7

AA001294  
ID AA001294 standard; DNA; 840 BP.

XX AA001294;

AC AA001294;

DT 12-OCT-2000 (first entry)

DE Ehrlichia canis immunoreactive protein Eca28SA3 DNA.  
KW Homologous 28-kDa protein gene; Eca28SA3; immunoreactive; vaccine;  
KW p28 gene; polymorphic multiple gene family; immunoprotective antigen;  
KW antibacterial; canine ehrlichiosis; canine tropical pancytopenia;  
KW tick-borne rickettsial disease; serodiagnosis; ds.

OS Ehrlichia canis.

XX Key Location/Qualifiers

PH CDS 1..840  
FT /\*tag= a  
FT /product= "Eca28SA3 protein (30-kDa)"  
FT /note= "Does not include stop codon"  
FT /partial  
FT sig\_peptide 1..69  
FT /\*tag= b  
FT mat\_peptide 70..840  
FT /\*tag= c  
FT /product= "Mature Eca28SA3 protein (28-kDa)"

WO200032745-A2.

08-JUN-2000.

\*-NOV-1999; 99WO-US28075.

\*-NOV-1999; 98US-0201458.

\*-NOV-1999; 99US-0261358.

RE-) RES DEV FOUND.

XX Walker DH, Yu X, McBride JW;  
PI WPI; 2000-412298/35.  
XX P-PSDB; AAY71479.  
XX Ehrlichia canis antigens useful for vaccinating against canine  
PT ehrlichiosis in dogs -  
XX Claim 5; Page 67-68; 86pp; English.  
XX The patent relates to homologous 28-kilodalton (kDa) protein genes of  
CC Ehrlichia canis, designated Eca28SA1, Eca28SA2, Eca28SA3, Eca28-1 and  
CC Eca28-2. These genes are members of a polymorphic multiple gene family  
CC and contained in a single locus of 5.592 kb. The 28-kDa proteins are  
CC immunoreactive with anti-E. canis serum hence are important  
CC immunoprotective antigens. The protein is useful for vaccinating  
CC against E. canis infections such as canine ehrlichiosis in dogs.  
CC Canine ehrlichiosis, also known as canine tropical pancytopenia, is a  
CC tick-borne rickettsial disease of dogs. Eca28-1 is conserved amongst  
CC different strains of E. canis and hence useful for serodiagnosis of  
CC canine ehrlichiosis. The present sequence is a DNA encoding E. canis  
CC Eca28SA3 30-kDa protein which is post-translationally modified to a  
CC mature 28-kDa protein by cleavage of N-terminal signal sequence.  
XX Sequence 840 BP; 282 A; 137 C; 150 G; 271 T; 0 other;  
SQ  
Query Match 27.9%; Score 234; DB 21; Length 840;  
Best Local Similarity 59.0%; Pred. No. 9.4e-46;  
Matches 503; Conservative 0; Mismatches 325; Indels 24; Gaps 5;  
QY 1 ATGAATTATAGAAAAATCTAGTAGAAGCGGTTAACTCAATTAATGTCAATCTTACCA 60  
Db 1 ATGAATTGCAAAAAAATCTTATAACAACATGCAATTAATGTCAATTAATGTCTTCCA 60  
QY 61 TATCAGTCTTTTGGCAGATCCTGTAGGTTCAAGAACATAATGATAACAAGAAAGCTTCTAC 120  
Db 61 AGCATATCTTTTCTGTACTATATACAAG-----ACGATPACACTGGTAGTCTTAC 111  
QY 121 ATTAGTGCAAAAGTACAATCCAAAGTATATCACATTTAGAAAAATCTCTCTGAAGA--AA 178  
Db 112 ATCAGTGGAAAAATATGTACCAAGTCTTTCACATTTTGGTGTCTTCTCAGCTAAAGAAGAA 171  
QY 179 CTCCTATTAAATGGAACAAATCTCTCACTAAAAAAGTTTCGGACTAAAGAAAGATGGTG 238  
Db 172 AGAAACTCAACTGTTGGAGTTTTCGATTAATAAACAATGATTTGAATGGAGGTACAATATCT 231  
QY 239 ATATAACAAAAAAGACGATTTTACAAGAGTAGTCCAGGCAATGATTTTC-AAAAATAAC 297  
Db 232 AACTCTTCTCCGAAAAATATATTACAGTTTCAAAATTTTCGTTAAATACGNAACAAC 291  
QY 298 TTAATATCAGGATTTTCAGGAAGTATTTGTTACTCTATGGACGGACCAAGATAGAACTT 357  
Db 292 CCATTTCTTAGGGTTTGCAGGAGCTATTGTTTATTTCAATGGGTGGCCCAAGATAGAACTT 351  
QY 358 GAAGCTGCATATCAACAATTTAATCCAAAAAACACCGATAACAATGATGATGATTAATGGT 417  
Db 352 GAAATTCTGTACGAGACATTCGATGTGAAAAATCAGAACAAATATATATAAGAACCGCGCA 411  
QY 418 GAATACTATAAACATTTTGTGCAATTAATCTGTAAGATGCAATG-----GAAGATCAG 468  
Db 412 CACGATACTGTGCTTTATCTCATCATAGTTCCAGCAACAGCATGTCCTCCGCAAGTAAC 471  
QY 469 CAATATGTAGTACTTTAAAAAATGACCGCATAACTTTTATGTCTCATGATGTTAATCTTGC 528  
Db 472 AAATTTGTTTCTTAAAAAATGAAGGTTAATTTGACTTATCATTTATGATAAATGCATGC 531  
QY 529 TATGACATTAAGCTGAAGGAGTATCTTTTCGATACCATATGATGTCAGGTATAGAGCA 588  
Db 532 TATGACATAAATAATTTGAAGGAATGCTTTTTCACCTTATATTTGTGAGGTGTGTACT 591  
QY 599 GATCTTATCACTATTTTAAAGACCTCAATCTTAAATTTGCTTTACCAAGAAAAATAGGT 648



Oy		61	TATCAGCTTTTGGAGATCCGTGAGGTTCAAGAACATAATGATTAACAAGAAAGGCTTCAC	120
Db		61	AGCATATCTTTTCTGATACTATACAG-----ACGATPACACTGGTAGCTTCTAC	111
Oy		121	ATTAGTCGAAGTACCAATCCAGATATACACACTTTAGAAAAATTCTCGTGAAG--AA	178
Db		112	ATCAGTGGAATAATGATACCAGATGTTTCACATTTTGCTGTTTTCTCAGCTTAAGAGNA	171
Oy		179	CTCCTATTAAATGGAACAATTTCTCACTAATAAAAAGTTTGGACATAAAGAAAGATGGG	238
Db		172	AGAAACTCACTGTGGAGTTTGGATTTAAACATGATGGAATGGAAGGTACCAATCT	231
Oy		239	ATATPACAAAAAAGACATTTTACAGAGTAGCTCCAGCAATTGATTTT-C-AAAAAAC	297
Db		232	AACCTCTCCAGAAAAATATATTACAGCTCCAAAATTATTCGTTTAATACGAAAAACAC	291
Oy		298	TTAAATACAGAGATTTTCAGAGATATGGTACTCTATGACGACCAAGAAATGACCTT	357
Db		292	CCAATCTTAGAGTTTGACGAGCATATGGTATTATCAATGGGGTGCCCAAGATPAGACTT	351
Oy		358	GAACTGCATATCAACAATTTAATCCAAAAACCCGATAACAATGATACTGATATAGT	417
Db		352	GAACTTCGTGCGAGACACTGATGATGAAAAATCAGAACATTAATTAAGAACGCCGA	411
Oy		418	GAATACATTAACAATTTTGCAATATCTCGTAAAGATGCAATG-----GAAGTACG	468
Db		412	CACGATATCTGTGCTTATCTCATATAGTTTCAGACAAACAGATCTCTCCGCCAGTAC	471
Oy		469	CAATATGATGACTTAATAAATGACGGCATTAATTATGTCAATGATGGTTAATACCTGC	528
Db		472	AAATTTGTTTTCTTAATAAATAGAAAGGTTAATGACTTATCATTTATGATTAATGATGC	531
Oy		529	TATGACATTACAGCTGAAGAGATCTTTCGTAACATATGCATGTGCAAGTATAGAGCA	588
Db		532	TATGACATAATTAATGGAAGATGCTTTTTCACCTTAATTTGTGCAAGGTGTGTACT	591
Oy		589	GATCTATACATATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGAAAAATAGGT	648
Db		592	GATGTGTTTCAATGTTGAAGCTATTAATCTTAATTTCTTACCAAGAAAACTAGGA	651
Oy		649	ATTAGTACCCTATCACACAGAACTCTCTGCAATTATTTGGTAGTACTACATGCGCTT	708
Db		652	TTAGTATATAGTATATAGTTTCAAGAGCCTCTGTTTTTATCCGGTGAACATTTACAGAGTIC	711
Oy		709	ATTGCTAATTAATTGGAGAAGATACCTGTATATACCTCTGTATGATTAATGATGCTCCT	768
Db		712	ATAGGATATGATTTAGAGACATCCCTGTATGTTCCATGATGATCAA---ATCTTCCA	768
Oy		769	CAAACCAATCTGCTTCACTAATCTTGACGTTGATACCTTTGGCGGAGAAATGGAATG	828
Db		769	GAAAAACAATTTGCATATGATACACTTAATATGTGTCTCACTTTGGCATPAGAACTTGGAGGA	828
Oy		829	AGGTTACCTTC 840	
Db		829	AGATTTAACCTTC 840	
<hr/>				
RESULT 9				
ADD01295				
ID	ADD01295	standard; DNA; 2037 bp.		
XX	ADD01295;			
XX				
DT	12-OCT-2000	(first entry)		
XX				
DE	Ehrlichia canis immunoreactive protein genes Eca28SA2 and Eca28SA3.			
XX				
KW	Homologous mature 28-kDa protein gene; Eca28SA2; Eca28SA3; vaccine;			
KW	p28 gene; polymorphic multiple gene family; immunoprotective antigen;			
XX	antibacterial; canine ehrlichiosis; canine tropical pancytopenia;			
XX	tick-borne rickettsial disease; serodiagnosis; immunoreactive; de.			
OS	Ehrlichia canis.			

Key	Location/Qualifiers
CD5	1..852
FT	/*tag= a
FT	/product= "Eca28SA2 protein (30-kDa) "
FT	853..1194
FT	/*tag= b
FT	/note= "intergenic non-coding region NC2 "
FT	1195..2037
FT	/*tag= c
FT	/product= "Eca28SA3 protein (30-kDa) "
FT	1195..1263
FT	/*tag= d
FT	1264..2034
FT	/*tag= e
FT	/product= "Mature Eca28SA3 protein (28-kDa) "
XX	
XX	WO200032745-A2.
XX	
XX	08-JUN-2000.
XX	
XX	24-NOV-1999; 99MO-US28075.
XX	
XX	30-NOV-1998; 98US-0201458.
XX	
XX	03-MAR-1999; 99US-0261358.
XX	
XX	(RERE-) RES DEV FOUND.
XX	
XX	Walker DH, Yu X, McBride JW;
XX	
XX	WPI; 2000-412298/35.
XX	
XX	P-PSDB; AAY71478, AAY71479.
XX	
XX	Ehrlichia canis antigens useful for vaccinating against canine
XX	ehrlichiosis in dogs -
XX	
XX	Example 15; Fig 7; 86pp; English.
XX	
XX	The patent relates to homologous 28-kilodalton (kDa) protein genes of
XX	Ehrlichia canis, designated Eca28SA1, Eca28SA2, Eca28SA3, Eca28-1 and
XX	Eca28-2. These genes are members of a polymorphic multiple gene family
XX	and contained in a single locus of 5.592 kb. The 28-kDa proteins are
XX	immunoreactive with anti-E. canis serum hence are important
XX	immunoprotective antigens. The protein is useful for vaccinating
XX	against E. canis infections such as canine ehrlichiosis in dogs.
XX	Canine ehrlichiosis, also known as canine tropical pancytopenia, is a
XX	tick-borne rickettsial disease of dogs. Eca28-1 is conserved amongst
XX	different strains of E. canis and hence useful for serodiagnosis of
XX	canine ehrlichiosis. The present sequence is that of homologous
XX	genes encoding E. canis Eca28SA2 and Eca28SA3 30-kDa proteins which
XX	are post-translationally modified to corresponding
XX	mature 28-kDa proteins by cleavage of N-terminal signal sequence.
XX	
XX	Sequence 2037 BP; 687 A; 321 C; 342 G; 687 T; 0 other;
XX	
XX	Try Match 27.9%; Score 234; DB 21; Length 2037;
XX	Local Similarity 59.0%; Pred. No. 1.1e-45;
XX	thes 503; Conservative 0; Mismatches 325; Indels 24; Gaps 5;
XX	
XX	1 ATGAATTAAAGAAATTCCTAGTAAAGCGGTTAACTCATTAATGCAATCTTACCA 60
XX	
XX	1195 ATGAATTGCAAAAATTCCTTAACAACTGCAATTAATGCAATTAATGCAATCTTCA 125
XX	
XX	61 TATCAGCTTTTGGAGATCTGTAGTTCAAGAACTAAATGATTAAGAAAGGCTTTTAC 120
XX	
XX	1255 AGCAATCTTTTTCGATACATACAG-----ACGATTAACACGTGAGCTTTTAC 1305
XX	
XX	121 ATTAGTGAAGTACAAATCAAGTATATACACTTTAGAAATTTCTTGCTGAAGA--AA 178
XX	
XX	1306 ATCAGTGAAGAAATATGTACCAAGTGTTCACATTTGGTGTTTCTCAGCTTAAGAAAGA 1365
XX	
XX	179 CTCCTATTAAAGCAAAATTTCTCCACTTAAGAAAGTTTTCGACTTAAGAAAGATGGTG 238
XX	

Db	1366	AGAAACTCACTGTTGGAGTITTTGGATTAAACATGATTGGAAATGGAGGTACATAACT	1425
Qy	239	ATATPACAAAAAAGAGATTTTACAAGAGTAGTCCAGGACTGATTTTC-AAAATTAC	297
Db	1426	AACCTTCTCCAGAAATATATTACAGTTCCAAATTAATTTGTTAAATACGAAACAC	1485
Qy	298	TTAATATCAGGATTTTTCAGGAAGTAGTGTACTCTATGACGACCAAGATAGACTT	357
Db	1486	CCATTCTTAGGGTTTGGACGAGCATTTGGTTATTCATTTGGGTGGCCCAAGATAGACTT	1545
Qy	358	GAACTGCAATTCACAATTTAATCCAAAAACCCGATAACAATGATCTGATTAATGCT	417
Db	1546	GAACTTCGTACGAGCATTCGATGAAAAATGAAACATTAATTAAGAACGGCCCA	1605
Qy	418	GAATACATTAACAATTTTTCATTAATCTCTGTAAGATGCAATG-----GAAGATGAG	468
Db	1606	CACAGATACGTGCTTATCTCATCATAGTTCACCAACAGCATGCTCTCCGAAATAC	1665
Qy	469	CAATATGATGACTTAATAAATGACGGCATTACTTTATGTCATTTGATGATGTTAATACTTG	528
Db	1666	AAATTTGTTTCTTAAAAAATGAAAGGTAAATGACTTAATCAATTAATGATTAATGATGC	1725
Qy	529	TATGACATTCACGTGGAAGAGTATCTTTCGTACCATATGCAATGCAATGCAAGTATAGACCA	588
Db	1726	TATGACATTAATTAATGGAAGAAATGCTTTTCACTTAATTAATTTGTGCAAGGTGTGTACT	1785
Qy	589	GATCTTAATCACTATTTTAAAGACCTCAATCTAAATTTGCTTACCAAGAAAAATAGT	648
Db	1786	GATGTGTGTTCCAGTGTGGAAGCTATTAATCTTAATTTCTTTACCAAGGAAAACTGGA	1845
Qy	649	ATTAGTACCTTATCACACCGAAGTCTGTCATTTATTTGGTATGACTACATGAGCGTT	708
Db	1846	TTAGATTATAGATTAAGTTCGAAAGCTCTGTGTTTATCGGTGACACTTTCACAGATC	1905
Qy	709	ATTGATTAATAATTGGAAGATACCTGTATAATCTCTGTAGTATTAATAGATGCTCT	768
Db	1906	ATAGATATGAATTTAGAGACATCCCTGATAGTTCCTAGTATGATCAAA--ATCTTCCA	1962
Qy	769	CAAAACCAATCTGCTTCACTAATCTTGAAGTGTGATACCTTTGGCGAGAAATGGAATG	828
Db	1963	GAAACCAATTTGGCAATAGTAACTAAATGTGTGTCATTTGGCATTAAGAACTTGGAGGA	2022
Qy	829	AGGTTCACCTTC 840	
Db	2023	AGATTTAACCTTC 2034	
RESULT 10			
ID	AAK34762	AAK34762 standard; DNA; 843 BP.	
AC	AAK34762;		
XX			
DT	05-JUL-1999	(first entry)	
XX			
DE		DNA encoding P30-2 protein.	
XX			
KW		Outer membrane protein; OMP; Ehrlichia chafeensis; E. canis; P30;	
XX		detection; dog; ss.	
OS		Ehrlichia canis.	
PN	W09913720-A1.		
PD	25-MAR-1999.		
XX			
PF	18-SEP-1998;	98MO-US19600.	
XX			
PR	19-SEP-1997;	97US-0059353.	
XX			
PA	(OHIS )	UNIV OHIO STATE.	
XX			
PI	Ohashi N, Rikihisa Y,		

XX WPI; 1999-254290/21.  
 XX P-PSDB; AAY06962.  
 XX Novel outer membrane proteins from Ehrlichia chaffeensis and  
 XX Ehrlichia canis  
 XX Disclosure; Fig 22A; 55pp; English.  
 XX The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and  
 CC consist of proteins shown in AAY06959-970. The proteins and genes are  
 CC used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX Sequence 846 BP; 283 A; 136 C; 150 G; 274 T; 0 other;  
 SQ

Query Match 27.7%; Score 232.4; DB 20; Length 843;  
 Best Local Similarity 58.9%; Pred. No. 2.2e-45;  
 Matches 502; Conservative 0; Mismatches 326; Indels 24; Gaps 5;

QY 1 ATGAATTAATAGAAAATTTCTAGTAAGAGCGGTTAATCTCAATTAATGTCAATCTTACCA 60  
 DB 1 ATGAATTAATAGAAAATTTCTAGTAAGAGCGGTTAATCTCAATTAATGTCAATCTTACCA 60  
 QY 61 TATCAGTCTTTTGGAGATCCTGTAGGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120  
 DB 61 AGCATATCTTTTCTGATACTATACAAG-----ACGATAACACTGGTAGCTTCTAC 111  
 QY 121 ATTAGTGCAAGTACAATCCCAAGTATATACACTTTAGAAAATTTCTCTGCT---GAAGAA 177  
 DB 112 ATCAGTGGAAAATACATGCCAAGTCTTCCGATTTTGGAGTATTTCTCTGCTAAGGAAGA 171  
 QY 178 ACTCTTATTAATGAGGTTTTCGACTAAAGGTTTTCGACTAAGAAAGATGGT 237  
 DB 172 AGAAATACACACAGTTGGAGTGT.TGGACTGAAGCAAAATTTGGACGGAAGCGCAATATCC 231

QY 769 CAACACCATCTGCTTCACTAATCTTTCAGCTTGGATACATTTGGCGGAGAAATTCGATG 828  
 DB 769 GAAACCAATTTGCAATAGTAACACTAATGTGTCTACCTTTGGTTAGAACTTGGAGGA 828  
 QY 829 AGGTTCACTTTC 840  
 DB 829 AGATTAACTTC 840

RESULT 11  
 AAX34743  
 ID AAX34743 standard; DNA; 846 BP.  
 XX AC  
 XX AAX34743;  
 XX  
 DT 05-JUL-1999 (first entry)  
 XX  
 DE DNA encoding OMP-1 protein.  
 XX  
 KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 KW detection; dog; ss.  
 XX  
 OS Ehrlichia chaffeensis.  
 XX  
 FN WO9913720-A1.  
 XX  
 PD 25-MAR-1999.  
 XX  
 PP 18-SEP-1998; 98WO-US19600.  
 XX  
 PR 19-SEP-1997; 97US-0059353.  
 XX  
 PA (OHIS) UNIV OHIO STATE.  
 XX  
 PI Ohashi N, Rikihisa Y;  
 XX  
 DR WPI; 1999-254290/21.  
 DR P-PSDB; AAY06943.  
 XX  
 PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
 PT Ehrlichia canis  
 XX  
 PS Disclosure; Fig 3A; 55pp; English.  
 CC  
 CC The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and  
 CC consist of proteins shown in AAY06959-970. The proteins and genes are  
 CC used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX  
 SQ Sequence 846 BP; 281 A; 143 C; 164 G; 258 T; 0 other;  
 SQ

Query Match 27.7%; Score 232.4; DB 20; Length 846;  
 Best Local Similarity 58.1%; Pred. No. 2.2e-45;  
 Matches 495; Conservative 0; Mismatches 336; Indels 21; Gaps 4;

QY 1 ATGAATTAATAGAAAATTTCTAGTAAGAGCGGTTAATCTCAATTAATGTCAATCTTACCA 60  
 DB 1 ATGAATTAATAGAAAATTTCTAGTAAGAGCGGTTAATCTCAATTAATGTCAATCTTACCT 60  
 QY 61 TATCAGTCTTTTGGAGATCCTGTAGGTTCAAGAACTAATGATAACAAGAGGCTTCTAC 120  
 DB 61 GGAGTATCATTTTCGACCCAGCAGGTAGTGTGTTAATACGGTAAT-----TTCTAC 111  
 QY 121 ATTAGTGCAAGTACAATCCCAAGTATATACACTTTAGAAAATTTCTCTGCT---GAAGAA 177  
 DB 112 ATCAGTGGAAAATACATGCCAAGTCTTCCGATTTTGGAGTATTTCTCTGCTAAGGAAGA 171  
 QY 178 ACTCTTATTAATGAGGTTTTCGACTAAAGGTTTTCGACTAAGAAAGATGGT 237  
 DB 172 AGAAATACACACAGTTGGAGTGT.TGGACTGAAGCAAAATTTGGACGGAAGCGCAATATCC 231

QY 238 GATATAACAAAAAGACGATTTTACAGAGTAGCTCCAGGCAATGATTTTCAAAATAC 297  
 Db 232 AACTCCTCCCAACGATGATTTCACTGCTCABATTTATCATTTAATATGAACACAC 291  
 QY 298 TTAATATCAGATTTTTCAGGAAGATTTGGTACTCTATGAGCGACCAAGATGAACTT 357  
 Db 292 CCGTTTTTAGGTTTTTGCAGAGCTATTTGTTACTCAATGATGTTGTCGAAGATAGAGCTT 351  
 QY 358 GAAGCTGCATACCAATTTAAATCCAAAAACACCGATACCAATGATGATAAGGT 417  
 Db 352 GAAGTCTTATGAACATTTTGATGTAAATCAAGGTAAACAATTAAGAATGAAGCA 411  
 QY 418 GAATACATATAACATTTTGGCAATATCTGTAAGATCAATG-----GAAGATCAGCA 471  
 Db 412 CATAGATATTGTCTCTATCCCAACTCAGCAGCAGATGAGTAGTGAAGTAATAT 471  
 QY 472 TATGTAGTACTTAAATATGAGCGATACCTTTTATGTCATTTATGTTATATCTGCTAT 531  
 Db 472 TTTGTCTTTCTAAAAATGAAGGATTTACTTGACATATCATTTATGCTGAACGATGCTAT 531  
 QY 532 GACATTACAGCTGAAGGATATCTTTGATACCATATGATGTCAGGTATAGGAGAT 591  
 Db 532 GACGTAGTAGGAGGACATACCTTTTCTTATATATGCGCAGGTATCGGTACTGAT 591  
 QY 592 CTTATCACTATTTTAAAGACCTCAATCTAAAAATTTGCTTACCAGGAAAAATAGGTATT 651  
 Db 592 TTAGTATCCATGTTTGAAGCTACAAATCTTAAATTTCTTACCAGGAAATAGGTATT 651  
 QY 652 AGTTACCTTATCACACGAGGCTCTGCAATTTATTTGGTGAATACATACGCGGTATT 711  
 Db 652 AGCTACTCTATAGCCGAGGCTCTGTTTATTTGGTGGGCACTTTTATAGGTAAATA 711  
 QY 712 GGTAAATAATTTGAGAGATAC---CTGTAATAACTCTCTAGTATTAAATGATGCTCT 768  
 Db 712 GGAACGATTTAGAGATTTCTCTATATATACCTACTGGATCAACACTTGCAGGAAA 771  
 QY 769 CAACACCATCTCTTCACTGATCTTGAAGTTGGATATCTTGGCGGAGAAAATGGAATG 828  
 Db 772 GGAATACCTCTCAATAGTAACTGATGATGATGATGATGATGATGATGATGATGATG 831  
 QY 829 AGGTTCACTTC 840  
 Db 832 AGGTTGATTC 843

## RESULT 12

AAC68706

AAC68706 standard; DNA; 830 BP.

XX AC AAC68706;

XX DT 02-MAR-2001 (first entry)

XX DE Ehrlichia chaffeensis VSAS gene partial coding sequence.

XX KW Ehrlichia chaffeensis; VSAS, variable surface antigen 5; MAP1;

XX KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;

XX KW rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;

XX KW 3gdrf3; ds.

XX OS Ehrlichia chaffeensis.

XX PN W0200065063-A2.

XX XX 02-NOV-2000.

XX PF 21-APR-2000; 2000WO-US10886.

XX PR 22-APR-1999; 99US-0130725.

XX XX (UFL) UNIV FLORIDA.

XX PA

XX XX

PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;  
 PI Rurangitwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;  
 DR WPI; 2000-679675/66.  
 DR P-PSDB; AAB36189.

XX New polynucleotides useful as DNA vaccines for conferring immunity to  
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium  
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens

PS Claim 4; Page 41-42; 63pp; English.

XX The present sequence shows a high degree of similarity to the major  
 CC antigenic protein 1 (MAP1) gene of Ehrlichia sp. The MAP1 gene may be  
 CC used in a vaccine to protect animals or humans against rickettsial  
 CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,  
 CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response  
 CC protective against the rickettsial pathogen. The nucleic acid vaccines  
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.  
 CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1  
 CC and 3gdrf3 may be used in therapeutic and diagnostic applications. The  
 CC polypeptides are useful for detecting antibodies associated with  
 CC infection by a rickettsial pathogen whilst the polynucleotides may be  
 CC used to detect the presence of rickettsial nucleic acids.

XX SQ Sequence 830 BP; 278 A; 141 C; 160 G; 251 T; 0 other;

Query Match 27.2%; Score 228.6; DB 21; Length 830;

Best Local Similarity 58.2%; Pred. No. 1.8e-44;

Matches 487; Conservative 0; Mismatches 329; Indels 21; Gaps 4;

QY 1 ATGATATATAGAAATCTAGTAAGAGCGGTTATCTCATTAATGCTCAATCTTACCA 60

Db 3 ATGATATATAGAAATCTAGTAAGAGCGGTTATCTCATTAATGCTCAATCTTACCA 62

QY 61 TATCAGTCTTTTGCAGATCTCTGATAGTTCAGAACTAATGATAACAAAGAGGCTTCTAC 120

Db 63 GGAGTATCAATTTCCGACCCAGAGGTAGTGGTATTAACGGTAAAT-----TTCTAC 113

QY 121 ATTAGTCAAGATCAATCCAAAGTATATCACTTTTAGAAAAATTTCTCTGCT---GAAGAA 177

Db 114 ATCAGTGGAAAAATACATGCCAAGTCTTCCGATTTTGGAGTATTTCTCTGCTAAGAGAA 173

QY 178 ACTCCTATTATGGAACAAATTTCTCACTAAAAAGTTTTCGACATAAAGAAAGATGGT 237

Db 174 AGAATACAAAGTGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 233

QY 238 GATATAACAAAAAGACGATTTTACAGAGTAGCTCCAGGCAATGATTTTCAAAATAC 297

Db 234 AACTCCTCCCAACGATGATTTTCACTGCTCAATTTATTTAATATGAACAAAC 293

QY 298 TTAATATCAGATTTTTCAGGAAGTATTTGGTACTCTATGACCGACCAAGATAGAACTT 357

Db 294 CCGTTTTTAGGTTTTTGCAGGAGCTATTTGGTACTCTCAATGATGGTCCCAAGATAGAGCTT 353

QY 358 GAAGTGCATATCAACATTTTATCCAAAAACACCGATACCAATGATGATGATGATGATG 417

Db 354 GAAGTATCTTATGAACATTTTATGATGATGATGATGATGATGATGATGATGATGATG 413

QY 418 GAATACATATAACATTTTGGCAATATCTGTAAGATCAATG-----GAAGATCAGCA 471

Db 414 CATAGATATTGTCTCTATCCCAACTCAGCAGCAGATGAGTAGTGAAGTAATAT 473

QY 472 TATGTAGTACTTAAATATGAGCGATATCTTTTATGTCATTTGATGTTTAACTTTGCTAT 531

Db 474 TTTGTCTTTCTAAAAATGAAGGATTTACTTGACATATCATTTATGCTGAACGATGCTAT 533

QY 532 GACATTACAGCTGAAGGATATCTTTTCGTACCATATGATGTCAGGTATAGGAGCAGAT 591

Db 534 GACGTAGTAGGAGGAGCAGTACCTTTTCTCTTATATGCGCAGGTATCGGTACTGAT 593

QY 592 CTTATCACTATTTTAAAGACCTCAATCTTAAATTTGCTTTACCAAGGAAAAATAGGTATT 651

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 3, 2003, 22:38:34 ; Search time 1981 Seconds  
(without alignments)  
4304.453 Million cell updates/sec

Title: US-10-062-624-40  
Perfect score: 1496  
Sequence: 1 MNVKLFTIINTVLVCLLSL.....AVATLNTGYGGEIGARLTF 293

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US10062624/runat\_30062003\_091229\_24302/app\_query.faeta\_1.455  
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=spect -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10062624 @CGN 1 1 1687 @runat\_30062003\_091229\_24302 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pt: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_or: \*  
21: em\_Or: \*  
22: em\_Ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*

29: em\_vi: \*  
30: em\_htg\_hum: \*  
31: em\_htg\_inv: \*  
32: em\_htg\_other: \*  
33: em\_htg\_mus: \*  
34: em\_htg\_pln: \*  
35: em\_htg\_rod: \*  
36: em\_htg\_mam: \*  
37: em\_htg\_vrt: \*  
38: em\_sy: \*  
39: em\_htgo\_hum: \*  
40: em\_htgo\_mus: \*  
41: em\_htgo\_other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1496	100.0	11329	1	AF082744 Ehrlichia
2	1496	100.0	28254	1	AF078553 Ehrlichia
3	1072	71.7	27190	1	ECU72291 Ehrlichia C
4	1033	69.1	14759	1	AF230642 Ehrlichia
5	494.5	33.1	6913	1	AF324792 Ehrlichia
6	484.5	32.4	3507	1	AF125276 Ehrlichia
7	484.5	32.4	3551	1	AF125275 Ehrlichia
8	479.5	32.1	1101	1	AF125274 Ehrlichia
9	479.5	32.1	3535	1	AF125274 Ehrlichia
10	479.5	32.1	3538	1	AF125277 Ehrlichia
11	479.5	32.1	3541	1	AF125279 Ehrlichia
12	479.5	32.1	3572	1	AF125278 Ehrlichia
13	476.5	31.9	863	1	AF35200 Ehrlichia
14	472.5	31.6	1278	1	AF35200 Ehrlichia
15	470.5	31.5	873	1	AF368001 Ehrlichia
16	469.5	31.4	1278	1	AF368001 Ehrlichia
17	468.5	31.3	843	6	AX042314 Sequence
18	468.5	31.3	4683	1	AF062761 Ehrlichia
19	467.5	31.2	1584	1	AF062761 Ehrlichia
c 20	464	31.0	2977	1	AF068234 Ehrlichia
c 21	464	31.0	27190	1	ECU72291 Ehrlichia C
22	461.5	30.8	873	1	AY028378 Ehrlichia
23	460	30.7	1265	1	AF368000 Ehrlichia
24	459.5	30.7	831	1	AF368000 Ehrlichia
25	459	30.7	864	6	AX042305 Sequence
26	456.5	30.5	840	1	AF368012 Ehrlichia
27	456.5	30.5	1467	1	CRMAP1
28	456	30.5	1263	1	AF368009 Ehrlichia
29	454.5	30.4	840	1	AF368009 Ehrlichia
30	452	30.2	1263	1	AF368009 Ehrlichia
31	451.5	30.2	1282	1	AF368009 Ehrlichia
32	451.5	30.2	1312	1	AF077734 Ehrlichia
33	451	30.1	1280	1	AF368009 Ehrlichia
34	449.5	30.0	1243	1	AF077733 Ehrlichia
35	449.5	30.0	1283	1	AF077733 Ehrlichia
36	449.5	30.0	1307	1	AF368009 Ehrlichia
37	449.5	30.0	1309	1	AF368009 Ehrlichia
38	449	30.0	825	1	AF368008 Ehrlichia
39	448.5	30.0	834	1	AF368010 Ehrlichia
40	447.5	29.9	1286	1	AF368010 Ehrlichia
41	445.5	29.8	828	1	AF368014 Ehrlichia
42	444.5	29.7	828	1	AF368007 Ehrlichia
43	444	29.7	1315	1	AF368007 Ehrlichia
c 44	444	29.7	28254	1	AF078553 Ehrlichia
45	443.5	29.6	834	1	AF368011 Ehrlichia

ALIGNMENTS

RESULT 1



AGATCYSMGGRIPRIEVLVETFDVKNQNNYKNGAHRYCALSHSHSATSSASNKVF  
FLKNEGIDISFNACVYDIIIEGMPSPYICAGVGDVWSMFEAVNPKISYQCKLGL  
GYSISSEASVFIGGPHFRVIGNEFRDIPAMVPSSNLPENQFAIVTLNVCHFGLGEG  
RNFN"

gene

7349, .8185

CDS

/gene="p28-7"

7349, .8185

/gene="p28-7"

/note="outer membrane protein"

/codon\_start=1

/transl\_table=1

/product="p28-7"

/protein\_id="AAC64550.2"

/db\_xref="GI:10181082"

/translation="MNCKKILITALLISLMYSIPISFSDTIQDGNMGNFYISGKTV

PSVSHFGSFAKESKSTGVGFLKHDWDGSPILKNKHADPAFVNYSFRVNNPFLGF

AGATCYSMGGRIPRIEFLSYBAFDVKSPNINONDARHYCALSHSHSATSEADKFLFK

NEGIDISFNACVYDIIINDKVPSPYICAGIGTDLISMEFATSPKISYQCKLGISIS

INPTSVFIGGPHFRVIGNEFRDIPALVPSNITISGFPATVTLNVCHFGLGGRF

NF"

8538, .9404

/gene="p28-8"

8538, .9404

/gene="p28-8"

/note="outer membrane protein"

/codon\_start=1

/transl\_table=1

/product="p28-8"

/protein\_id="AAG14362.1"

/db\_xref="GI:10181090"

/translation="MNCKRFFIASALISLMSFLPSVSFSSIHEDNINNGNPFYISAKYM

PSASHFGVSFKENKTITGVFLKQDWDGATIKDASSHTIDPSTFISINYSFKYE

NNPFLGAGATCYSMGGRVFEVSEYEIFDVKNQNSYKNDHKYCALSHRHTGMPQA

GHQKFLVKNEGLLDLSLIMNACVYDITIDMPSPYICAGIGSDLVSMFETNPKIS

YQCKLGVSYSTSPSEASVFGGPHFRVIGNEFKDIPALTPAGATEIKGTQTTVTLNIC

HFGLGGRFTF"

complement(10031, .10846)

/gene="p28-9"

complement(10031, .10846)

/gene="p28-9"

/note="outer membrane protein"

/codon\_start=1

/transl\_table=1

/product="p28-9"

/protein\_id="AAG14363.1"

/db\_xref="GI:10181091"

/translation="MNCKRFFVGVLTSTFVFLSDGAFSDANFSEGRGLYIGSOYKV

GIPNFSNFAEETIPGTTKIFALGLDKSINTHSNTRSYDPTIASSPAGFSIGY

YVNDFRVEFGSYENFPERQWYSPNSQYKFPALSNATNSNKKFVILNENGVVDKS

LVNVYCDIASGSIPLAPYAGVGADYIKFLGISLPKFSYQVKFVYVNLNVNTMLF

GGGYHKKVGDHREVRVIAVHTALSVPRTTSASATLNTDYPGWEIGRPFAL"

BASE COUNT 4065 a 1825 c 1662 g 3777 t

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

1

Gaps:

0

US-10-062-624-40 (1-293) x AF082744 (1-11329)

QY

1 MethAenLysLeuLysPheThrIleLeuValLeuValCysLeuLeuSerLeu 20

Db

170 ATGAATATAAATCAAAATTTACTATAATAACACAGATTAGTCTTATTGTCATTA 229

QY

21 ProAenLysSerSerLysAlaIleAenAenAenAlaLysLysLysLysLysLysLys 40

Db

230 CCTATATATCTCTCTCAAGGCCATAACATACCGCTAAAGAGTACTACGGATTATAT 289

QY

41 IleSerGlyGlnTyrLysProSerValSerValPheSerAenPheSerValLysGluThr 60

Db 290 ATCAGTGGCAATATAAACCCAGTGTTCTGTTTTCAGTAATATTTTTCAGTAAGAACC 349

QY 61 AsnValIleThrLysAenLeuIleAlaLeuLysLysAenValAenSerIleGluThrLys 80

Db 350 AATGTCATAACTAAACCTTATAGCTTTTAAANAAGATGTTGACTCTATTATGAACCAAG 409

QY 81 ThrAspAlaSerValGlyIleSerAenProSerAenPheThrIleProTyrThrAlaVal 100

Db 410 ACTGATGCCAGTGTAGTATTAGTAACCCATCAATTTTACTATCCCTTACACAGCTGTA 469

QY 101 PheGlnAspAenSerValAenPheAenGlyThrIleGlyTyrThrPheAlaGluGlyThr 120

Db 470 TTTCAAGATAATCTGTCAATTTCAATGGAACTATTGGTTTACACCTTTTGTGAAGGTACA 529

QY 121 ArgValGluIleGluGlySerTyrGluGluPheAenValLysAenProGlyGlyTyrThr 140

Db 530 AGAGTTGAATAGAGGTTCTTATAGGAATTTGATGTTTAAANAACCTTGGAGGCTATACA 589

QY 141 LeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAenSerPheThr 160

Db 590 CTAAGTGCATGCCTATCGCTATTTTGCATTAGCACGTGAAATGAAAGCTAATAGTTTACA 649

QY 161 ProLysGluLysValSerAenSerIlePheHisThrValMetArgAenAenGlyLeuSer 180

Db 650 CCTAAAGAAAAGTTTCTAATAGTATTTTTCACACTGTAATGAGAAATGATGGATTATCT 709

QY 181 IleIleSerValIleValAenValCysTyrAspPheSerLeuAenAenLeuSerIleSer 200

Db 710 ATAAATCTCTGTTATAGTAATGTTTGTCTAGATTCTCTTTGAACAATTTGTCTAATATCG 769

QY 201 ProTyrIleCysGlyGlyAlaGlyValAlaIleGluPhePheAenValLeuHisIle 220

Db 770 CCTTACATATGTGGAGGAGCAGGGTAGATGCTATAGAAATCTTCGATGATTACACATT 829

QY 221 LysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAenIleSerLeu 240

Db 830 AAGTTTGCATATCAAGCAAGCTAGGTATTGCTTATTTCTCTACCATCTAACATTAGTCTC 889

QY 241 PheAlaSerLeuTyrTyrHisLysValMetGlyAenGlnPheLysAenLeuAenValGln 260

Db 890 TTTGCTAGTTTATATACCAATAAGTAATGGGCATCAATTTTAAATTTAAATGTCCAA 949

QY 261 HisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAenIle 280

Db 950 CATGTTGCTCAACTTGCAGTATACCTAAAATTTACATCCGAGTTGCTACACTTAATATT 1009

QY 281 GlyTyrPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293

Db 1010 GGTATTTTGGAGGTGAAATTTGGTGAAGATTGACATTT 1048

RESULT 2

AF078553 28254 bp DNA linear BCT 02-APR-2001

LOCUS Ehrlichia canis major outer membrane protein P30 multigene cluster

DEFINITION 1, complete sequence.

ACCESSION AF078553 AF078554 AF078555 AH006958

VERSION AF078553.2 GI:13512584

KEYWORDS

SOURCE Ehrlichia canis.

ORGANISM Ehrlichia canis

Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

Anaplasmataceae; Ehrlichia.

1 (bases 1 to 28254)

Chashi, N., Unver, A., Zhi, N. and Rikihisa, Y.

Cloning and characterization of multigenes encoding the

immunodominant 30-kilodalton major outer membrane proteins of

Ehrlichia canis and application of the recombinant protein for

serodiagnosis

J. Clin. Microbiol. 36 (9), 2671-2680 (1998)

98371112

MEDLINE 9705412

PUBMED

REFERENCE 2 (bases 1 to 28254)

AUTHORS Ohashi, N., Rikihisa, Y. and Unver, A.  
 TITLE Analysis of transcriptionally active gene clusters of major outer membrane protein multigene family in *Ehrlichia canis* and *E. chaffeensis*  
 JOURNAL Infect. Immun. 69 (4), 2083-2091 (2001)  
 MEDLINE 21153566  
 PUBMED 11254561  
 REFERENCE 3 (bases 1 to 28254)  
 AUTHORS Ohashi, N., Unver, A., Zhi, N. and Rikihisa, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-JUL-1998) Department of Veterinary Biosciences, The Ohio State University, 1925 Coffey Road, Columbus, OH 43210, USA  
 REFERENCE 4 (bases 1 to 28254)  
 AUTHORS Ohashi, N., Rikihisa, Y. and Unver, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-NOV-2000) Department of Veterinary Biosciences, The Ohio State University, 1925 Coffey Road, Columbus, OH 43210, USA  
 REMARK Sequence update by submitter  
 COMMENT On or before Apr 2, 2001 this sequence version replaced gi:3790556, gi:3790555, gi:3790558, gi:3790557.  
 FEATURES  
 source  
 1..28254  
 location/Qualifiers  
 /organism="Ehrlichia canis"  
 /strain="Oklaboma"  
 /db\_xref="taxon:944"  
 41..615  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical transcriptional regulator"  
 /protein\_id="AAK28679.1"  
 /db\_xref="GI:13512585"  
 /translation="PEYANEIKAHDPDIEDIEKNIQKHKTGEGIRLDVVDYASKNF  
 KKEGASEKLSATKVPHPVDECGVEKREIKRQIRMGSONQLANKIGITFQQQKY  
 EKGNTIRIVSLYQLASVINEVRDMLKQEDLNKISCDNPITPPHALRDNEEKFLP  
 BFNKSIDSKVELVMVRAVTCIKNEKVRNIYLVKALSLDNKS"  
 843..1733  
 /gene="p30-19"  
 /db\_xref="GI:13512586"  
 /translation="MKLYHLDNIMKFSAGIVSFIALFAPNAPSPVPIDFESNS  
 EMAGYASQYNIQGPFPSPISAKYTKDEKSEKTLFLSKEETETIDLKAGDFKKG  
 YSPVYRNVTGPGSAGIGYGGGLRVELEGSFTRFDVQKYNPDGHRVFAISKDSFI  
 QNSSSGSSNNKDYVMKNEGFAISLMFNACYDMIIGNSSLVNACIGIGQIIRFL  
 GGTNHTLPAKGLGFLISPKTILFANGYVYKADNFTNLSVQYPVEISNAPKHID  
 FIVYFNADNYCEVGLRFLF"  
 2491..3318  
 /gene="p30-18"  
 /db\_xref="GI:13512587"  
 /translation="WNNKLSLYIALILFTSHVSALVLDHNLVYFGIQYKPAHHLL  
 SNLLKRESSEVVEALDAIGSPLDSKEVNNFTIKINPHYDNNRLKGSVIFGYI  
 YKNPFRISSEIYELFOLKNEGKRGVFEKYPALFPAPSSQTQYDNNHLLNNGISIT  
 SALINACVDVLIPAHNIITYSLGFGIDIVDFLSKYITTKFSHQKGLGASYPISHRMSV  
 FTEVYHGLFGKGFQPLNLYNANTSPQPPHVHTTASAILSIGYGGSGVGIKFL"  
 3425..3871  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown function U2"  
 /protein\_id="AAK28682.1"  
 /db\_xref="GI:13512588"  
 /translation="WNNKVLQNEVLLALLPYSLHNSIDIEYNEQPIKLSRLNKNTI

gene  
 CDS  
 3882..4832  
 /gene="p30-17"  
 3882..4832  
 /gene="p30-17"  
 /note="P30 family member"  
 /codon\_start=1  
 /transl\_table=11  
 /product="major outer membrane protein P30-17"  
 /protein\_id="AAK28683.1"  
 /db\_xref="GI:13512589"  
 /translation="MNYAKVFLVCIILFPLSPSFAFNNNVFVHEIGKSGHGFYIGVO  
 YKCTPHNRFSTADSTFNLALSHKDYLFVSSTEVRLGESLPQSQNLHVGATGS  
 TLTUNTKDNKFPNGNPIYDNLGGVGGVGYISNNRIELEAFTEKRNKIKAPTGY  
 NYTDEYFAIATVYVYKTPVHVHKMGTGIISSFLNVTCTYDFTLKAKKIAPYLCL  
 GVSGDFDFLQOTRLKASQKAGLSYAIISPNLTFFVDGSGFHYGMNNQFPELLVDYPT  
 DISVMPSGDNATAYSEFTTMLAKLNMIFLAGSIGIRFIS"  
 4848..5699  
 /gene="p30-16"  
 4848..5699  
 /gene="p30-16"  
 /note="P30 family member"  
 /codon\_start=1  
 /transl\_table=11  
 /product="major outer membrane protein P30-16"  
 /protein\_id="AAK28684.1"  
 /db\_xref="GI:13512590"  
 /translation="MYKLYLSFTISLAQLLFGSFAFSDIKNNNHGSIYTIKYOPTI  
 SNKPNFKIETDFTDPIGDIAPNTDFDLKHNYSFVLYHKDSYKFYENDLSGL  
 ALSTGLAVKNIIRIEFGSKYKFDTKRLAYVHSREGHKFFAIPTSNFGVTPNEQNTYV  
 AKNGISILNINLCSETKEKNTFYICIGIGGDFIEI EDVMEVKETYOGKYGISVP  
 ITPKLVUSISGQYHKVIGNKFPLIQLPVALKRTDSDPKDQVLTALLTDLERFSSE  
 IGSFIF"  
 5716..6510  
 /gene="p30-15"  
 5716..6510  
 /gene="p30-15"  
 /note="P30 family member"  
 /codon\_start=1  
 /transl\_table=11  
 /product="major outer membrane protein P30-15"  
 /protein\_id="AAK28685.1"  
 /db\_xref="GI:13512591"  
 /translation="WYNIINIVYKTIALAFLLPRVSFSLIGNIEKSKLLSVHIN  
 SQKPSISQISNLYIQENSKKINILNLSNNTITTNMQLENTITNFRITIGFFKFR  
 LRFAVEDSYEFHDKDNLSKANLSKYSKYMNEDFQNTIATDNKLSITSAIVNICY  
 DILINNTTVLPILCTAVGICSTGFPDMRFKLYQKIGIGLYLINSNMLFPFNYYHK  
 VMRNKLNLTQYSVDINAFDAITVLANTDIGYFGSEVGVRFIEN"  
 6525..7394  
 /gene="p30-14"  
 6525..7394  
 /gene="p30-14"  
 /note="P30 family member"  
 /codon\_start=1  
 /transl\_table=11  
 /product="major outer membrane protein P30-14"  
 /protein\_id="AAK28686.1"  
 /db\_xref="GI:13512592"  
 /translation="MLQKLNFINLAFLLPFPQSFYLYIHDHEITQNVGLYISSQ  
 YKPSIPYKFNFLIENSHKTVLMGLANDVTHVTEYVLYKONTKENTPYSAKFRNSLIN  
 LSGAIGYSGQGPRLIEGSGYENFDVASCNKPVKANRYALVRDKKPGNIYPQDHS  
 HSNMSYTFYKNGGISILSVNWCYDIASNVKISPYVCAGIGGDFITLETWHIKF  
 AYQKFGISLYLSPSISIFANGHYHKVMDNVNFKNLHVKIYIKLQDAPITTSARAKLRI  
 GYFGSEYGVRFV"  
 7419..8255  
 /gene="p30-13"  
 7419..8255  
 /gene="p30-13"  
 /note="P30 family member"  
 /codon\_start=1  
 /transl\_table=11  
 /product="major outer membrane protein P30-13"



```
/protein_id="AAK28687.1"
/db_xref="GI:13512593"
/translation="MNNKSLIGITILLSPSSLPKAFSVINHSDISSNVNGLYFTG
QYRPAVSGFTVRENTIATQQLVSLNTNKNENHIERTNFSGIYTKAFODNAASF
GAIQSVPEGLKGFIEISYEKFGVKSTKNYSTNAVIFALAROTTSSNPNDKVVVK
NSGLSVASVMIINGCVNYSFYNLVSPVYICAGIGEDFLEFPDLVILKAYOGKLVYS
LSSRFNIFADMYHKVIGNQFNLVNVAVALDTPFKVTSALINLVAYFGGEVGIRF
IL"
```

CDS

```
8266..8823
/codon_start=1
/transl_table=11
/product="unknown function U3"
/protein_id="AAK28688.1"
/db_xref="GI:13512594"
/translation="MATYMIHAKIIEPMFSPSILTDYIILSLNNVDHNSFIICIT
KLVKILSSITKLVINLNTAVLHNLINLYLNKSKFKSTQSFYMFYFTKIKDR
ILNFNSLYFIKYNVYTKNSNESYKSTYKSHNFNATVTSFIIHQHKLKLYSITNQH
NKSTKSPSKILAKNILLNNQNSYE"
8816..9685
/gene="p30-12"
8816..9685
/gene="p30-12"
/note="P30 family member"
/codon_start=1
/transl_table=11
/product="major outer membrane protein P30-12"
/protein_id="AAK28689.1"
/db_xref="GI:13512595"
```

gene

CDS

```
/translation="MNSKTPYILGSLICLAACLPQSSSESNVNTYTKHTGLYIS
GLPKPSVSHSDPSIKETVNTALFGLKODISSILNKETQYNNFNVPYAKFQD
DPASFSVANGVIANGPRIIEGSEYEDFVNKPGNYTIDAHRYTLARKEISYKFA
PREKNYIIKNNGISIISIIINGCYDLSNDSKVSPYICTFGFGDFTEFSAIRFKFA
YQKIGISYLSLSNIIILTDGYHKVINSQFNLVNVEHVNELTDPKVTSAFALNI
EYFGGEFGLKFPF"
```

## Alignment Scores:

Pred. No.:	4,06e-116	Length:	28254
Score:	1496.00	Matches:	293
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-10-062-624-40 (1-293) x AF078553 (1-28254)

Qy	1	MetAenLysLeuLysPheThrIleAenThrValLeuValCysLeuLeuSerLeu	20
Db	14133	ATGATAATAACTCAATTTACTATAATAACACAGTATTAGTATGCTTATGCTATTA	14192
Qy	21	ProAenLysSerSerLysAlaLeAenAenAenAlaLysLysLysLysLysLysLys	40
Db	14193	CCTAATATATCTCTCAAGGCCATAAACAATAACGCCTAAAGAGTACTACGGATTATAT	14252
Qy	41	IleSerGlyGlnTyLysProSerValSerValPheSerAenPheSerValLysGluThr	60
Db	14253	ATCAGTGACAATAATAACCCAGTGTTCTGTTTCAGTAAATTTTCAGTTAAAGAACCC	14312
Qy	61	AsnValIleThrLysAenLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys	80
Db	14313	AATGTCATACTAAACCTTATAGCTTTAAAAAGAGTGTGACTCTATTGAAACCAAG	14372
Qy	81	ThrAenLysValGlyIleSerAenProSerAenPheThrIleProTyThrAlaVal	100
Db	14373	ACTGATGACAGTAGGTATTAGTAACCCATCAATTTTACTATCCCTATACAGCTGTA	14432
Qy	101	PheGlnAenSerValAenPheAenGlyThrIleGlyTyThrPheAlaGluGlyThr	120
Db	14433	TTTCAAGATAATCTGCAATTTCAATGGAAGTATTGTTACACCTTTGCTGAAGGTACA	14492
Qy	121	ArgValGluIleGluGlySerTyLysGluGluPheAspValLysAenProGlyGlyTyThr	140
Db	14493	AGAGTTGAAATAGAGGTTCTTATGAGGAATTTGATGTTAAACCCCTGGAGGCTATACA	14552

Qy	141	LeuSerAspAlaTyArgTyPheAlaLeuAlaArgGluMetLysGlyAenSerPheThr	160
Db	14553	CTAAGTGATCGTATCGTATTTTGCATTAGCAGCGTAAAGGTAATAGTTTAC	14612
Qy	161	ProLysGluLysValSerAenSerIlePheHisThrValMetArgAenAspGlyLeuSer	180
Db	14613	CCTAAGAAAAGTTTCTAATAGTATTTTCCACTGTAATGAGAAATGATGGATTATCT	14672
Qy	181	IleIleSerValIleValAenValCysTyAspPheSerLeuAenAenLeuSerIleSer	200
Db	14673	ATAATATCTGTATAGTAAATGTTTGTCTACGATTTCTTTTGAACAATTTGTCAATATCG	14732
Qy	201	ProTyIleCysGlyGlyAlaGlyValAspAlaIleGluPheAspValLeuHisIle	220
Db	14733	CCTTACATATGTCAGGAGCAGGAGTAGATGCTATAGATTTCTTCGATGATTACACATT	14792
Qy	221	LysPheAlaTyGlnSerLysLeuGlyIleAlaTySerLeuProSerAenIleSerLeu	240
Db	14793	AGTTTGCATATCAAGCAAGCTAGGTATTGCTTATTTCTTACCATCTAACATTAGTCTC	14852
Qy	241	PheAlaSerLeuTyTyHisLysValMetGlyAenGlnPheLysAenLeuAenValGln	260
Db	14853	TTTGTAGTTTATATACCATAAAGTAATGCGCAATCAATTTAAAAATTTAAATGTCCAA	14912
Qy	261	HisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAenIle	280
Db	14913	CATGTTGCTGAATGTCAGTATACCTAAATTTACATCCGAGTGTGTACACTTAATATT	14972
Qy	281	GlyTyPheGlyGlyGluIleGlyAlaArgLeuThrPhe	293
Db	14973	GGTATTTTGAGGTGAATTTGGTGAAGATTGACATT 15011	

RESULT 3

ECU72291	27190 bp	DNA	linear	BCT 02-APR-2001
LOCUS				
DEFINITION	Ehrlichia chaffeensis strain Arkansas major outer membrane protein			
ACCESSION	Omp-1 multigene cluster, complete sequence.			
VERSION	U72291 AF021338			
KEYWORDS	U72291.2 GI:13511827			
SOURCE	Ehrlichia chaffeensis.			
ORGANISM	Ehrlichia chaffeensis			
REFERENCE	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;			
AUTHORS	Anaplasmataceae; Ehrlichia.			
TITLE	1 (bases 14844 to 21136; 21479 to 22234)			
JOURNAL	Ohashi, N., Zhi, N., Zhang, Y. and Rikihisa, Y.			
MEDLINE	Immunodominant major outer membrane proteins of Ehrlichia			
PUBMED	chaffeensis are encoded by a polymorphic multigene family			
REFERENCE	Infect. Immun. 66 (1), 132-139 (1998)			
AUTHORS	98084465			
TITLE	2 (bases 1 to 27190)			
JOURNAL	Ohashi, N., Rikihisa, Y. and Unver, A.			
MEDLINE	Analysis of transcriptionally active gene clusters of major outer			
PUBMED	membrane protein multigene family in Ehrlichia canis and E.			
REFERENCE	chaffeensis			
AUTHORS	Infect. Immun. 69 (4), 2083-2091 (2001)			
TITLE	3 (bases 1 to 27190)			
JOURNAL	Ohashi, N., Zhi, N., Zhang, Y. and Rikihisa, Y.			
MEDLINE	Direct Submission			
PUBMED	Submitted (24-SEP-1996) Department of Veterinary Biosciences, The			
REFERENCE	Ohio State University, 1925 Coffey Rd., Columbus, OH 43210-1093,			
AUTHORS	USA			
TITLE	4 (bases 1 to 27190)			
JOURNAL	Ohashi, N., Zhi, N., Zhang, Y. and Rikihisa, Y.			
MEDLINE	Direct Submission			
PUBMED	Submitted (27-AUG-1997) Department of Veterinary Biosciences, The			
REFERENCE	Ohio State University, 1925 Coffey Rd., Columbus, OH 43210-1093,			
AUTHORS	USA			
TITLE	5 (bases 1 to 27190)			
JOURNAL	Ohashi, N., Rikihisa, Y. and Unver, A.			
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				

TITLE Direct Submission  
JOURNAL Submitted (29-NOV-2000) Department of Veterinary Biosciences, The Ohio State University, 1925 Coffey Rd., Columbus, OH 43210-1093, USA

REMARK  
COMMENT Sequence update by submitter  
On or before Apr 2, 2001 this sequence version replaced gi:2853584, gi:2853273.

FEATURES  
source Location/Qualifiers  
1..27190  
/organism="Ehrlichia chaffeensis"  
/strain="Arkansas"  
/db\_xref="taxon:945"  
315..956  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical transcriptional regulator"  
/protein\_id="AAC02808.2"  
/db\_xref="GI:13511828"  
/translation="MSTHAKNKNMEDTDEIKADHPLIAGLIDNMOKNPTNKGVN  
DDHSTKNTKSGVESTKDKNSTTKIRPHVDECVGKEIKQRIMRGSQNLANKLG  
ITFQOVQKEGTNRIVISRLYQLASVNVVRDMLKQDLKDKNISCDHTVASTSTL  
KDNEEKFPEFHSDSKIDSQVLMVMVRYATCIRKNEKVRNIIYNLKAULSDNQ"  
1136..2032  
/gene="omp-1M"  
1136..2032  
/gene="omp-1M"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1M"  
/protein\_id="AAK28658.1"  
/db\_xref="GI:13511829"  
/translation="MILINMKLFYHLDMTKFSAIGIVLSLVTLPACNVFASPIPD  
FNSDETSAPYASGHYNTQFRPSISVKYTKDENTERKELTFLTKESDTTPTRKQKS  
EFNDEKSGPIYRNRYTGFGAIGYGGGLRVELEGATRFDDVKQKTDYNNRIFAL  
CKKDSIESTDSNGHVMKNEGFRVISLTFNACYDMIVSNSSLVPSACIGIGQOITN  
FLGGNIHTLPKALGLGLISPKTVIFAPKVIYVKTQKNSFTNLSVQVPLEKEAPKH  
IDPIACFADNRYGGEVGLRFL"  
2498..3352  
/gene="omp-1N"  
2498..3352  
/gene="omp-1N"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1N"  
/protein\_id="AAK28659.1"  
/db\_xref="GI:13511830"  
/translation="MSKRSNRKFLVWMLILFTPHISLASVLNDHNSMVGVQYKPAR  
OHLKLLIKESAANTVEYGLKKDLNLLTGDKDNTFNFKYNYENNRLPGSGIP  
GYNNKRISELSYETPHIKNGYKRIDCEKHPALAKEISGGSNNPANNKYVTLIN  
NGISLTSALINVCYDVGDKGNIIITYSCLGFGVDITDLSKYTHTKFYSQGLGASIV  
SPQSVFIEGYVHGLFGKFEKIPVNPYCDYPSPTPPNSKPHVHTTALAMLSIGYGG  
SIGIKFRL"  
3459..3905  
/codon\_start=1  
/transl\_table=11  
/product="unknown function UN2"  
/protein\_id="AAK28660.1"  
/db\_xref="GI:13511831"  
/translation="WNNKSRGQVLLFAAYSIFKYINSCSTKFRLSVKFCFVNPLKYF  
YKSPFLKLGKFFVNYNYSKSNYLHRSILKFLGLIIRTNPKISVLFNFTPRVLIYY  
YKILKNKNSINYLNNKYINPTATSFINSVIGIYPGFVYNY"  
3915..4808  
/gene="omp-1Q"  
3915..4808  
/gene="omp-1Q"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1Q"  
/protein\_id="AAK28661.1"  
/db\_xref="GI:13511832"

gene  
CDS  
/translation="MSYAKVPILICILLVPSLSFAIVNNDFLKDNIGHFVIGQYKGP  
GVPRNPLVNNIRELMSSDECRSTIPHMVQSVAGQTLPEALEELAKGLHGGY  
LFTFLPNPYKKNLILGAGGVIGSYSTHFRVEAEAYEKENLTPAGYGLHKNFYEYFA  
LATMTDKHPHQAEDKYYKNTGTLTSPFIINACYDFLTKKTRNVAPYLCIGVGGN  
FIIDLDQVGFAYQAKVGSYFVSPNIAFFIDGSPFHGLNNOFSDLPVVDYSSSGFP  
TISAKFNANFLTSSIGIRFIS"  
4825..5682  
/gene="omp-1P"  
4825..5682  
/gene="omp-1P"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1P"  
/protein\_id="AAK28662.1"  
/db\_xref="GI:13511833"  
/translation="MQKLYISFILSGLLLPKPVFCMHQNNNDGVSVTIKYQJTTPH  
FKQFYKETDFDQEPITGLAKITANTFTDLTKENFSPLHQDTSYKSYQNDLIGLIGL  
SVGLFVKSPRIEPEGAYKNTKRLARYKSGDKGYFAI PKRSEHGLDNTFGYTVAK  
NNGLSIILSNILINLCSETKYKAPTPIYICIGVGDFLEIFDVMRIKFAYQGVGSYPIIT  
SKLILSINGQYHKVIGNKFFELPVYQPVVELKRLVTNKTSDKIDQDVTASLTNLNLEHS  
SEIGLSPIF"  
5695..6513  
/gene="omp-1T"  
5695..6513  
/gene="omp-1T"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1T"  
/protein\_id="AAK28663.1"  
/db\_xref="GI:13511834"  
/translation="MYNKKHYCYIVYVITLFFLLPIESLSALIGNVEKOLKVS  
TYVSSQKSLRIFHRNFSIQESHKPKSSEFKIKANLNILKSNALNLOFQDNTTSF  
SGTIGYSGKLRLAEAGCYQEFNKNNSNLSIISNNKYHRIHDEYAITNNKLSIA  
SMYNTCDISINNTSIVPYLCTGIGEDLVGLNTIHFKLAYQKGVGMSYLNNNILL  
FSDIYHYKVMGRNFKNLYQVADPNISEETIPI LAKLDIGYFGSEIGIRFMEN"  
6529..7416  
/gene="omp-1U"  
6529..7416  
/gene="omp-1U"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1U"  
/protein\_id="AAK28664.1"  
/db\_xref="GI:13511835"  
/translation="MTKKFNVNVILTFLLFPLKSFYTYANNNTITQKVLVYISQ  
YKPSIPHKFNSVEENDKVDLIGLTDVTYITEHLRDNTKFNTHYIAKPKNNFINP  
SSAIGYVSGGPRLEIESYGDVNVKNVAVQDVNRYPALVREKNGSNFSPKPHET  
SQPDSNPKSFYTLKMGNGVFVASVILNGCYDSENNNTISPYVCIGVGDFIEFE  
VMIKPCACQKVGISYFISFISITIFADAHYHKVNNKFNLLHVKISYELKNSPITSA  
TAKLNIEYFGGVEGMRFI"  
7438..8277  
/gene="omp-1V"  
7438..8277  
/gene="omp-1V"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1V"  
/protein\_id="AAK28665.1"  
/db\_xref="GI:13511836"  
/translation="MSKKFITGTCTVASLLSFLSIESFSAINHNHTGNTSGIYITG  
QYRFGVSHFNFSVNETNVDITQLVGYKKSASSIDPNTYFNQGPYITVFDONASFS  
GAIGYSYPSLRLEGSYFKDVKDYSKADAFRFPALARTSTVDPQAQYTM  
KNGLSVASIMINGCYDLNFNNLVSPYICAGIGEDIEFDLTHIKLAYQGLGISY  
YFRPKINVAGGYVHRVIGNKFNLVNVTLDLDFPKATSATVATLVAVYFGGAGVK  
FTF"  
8300..8890  
/codon\_start=1  
/transl\_table=11  
/product="unknown function UN3"



```

GLGKKKBEKIPVNYPCDYPSPPPNPKPHVHTALAMLSIGYGGSIGIKFIL"
3158. .3604
/translation="component of 28-kDa outer membrane protein
multigene family"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF73409.1"
/db_xref="GI:8132826"
/translation="MNNKSRGQVLLPAAYSIKYINCTSKFRLSKCFVYNPLKYF
YKSTFLKGLFFVNYNYSKTSNYLHSLFKLFGIIRTNFKISYLNFTNPRVLIYY
YKILNKNSSINLYLNKYNTINPTATSPINSIIGIYVGFVNNY"
3614. .4507
/gene="p28-2"
3614. .4507
/gene="p28-2"
/translation="component of 28-kDa outer membrane protein
multigene family"
/codon_start=1
/transl_table=11
/product="p28-2"
/protein_id="AAF73410.1"
/db_xref="GI:8132827"
/translation="MSYAKVFILICILLVPSLSFAIVNNDPLKDNIGHYVIGQYKP
GVPRNRELVTNNIRLMSSDECRSTIHPMVQSAOGLTPPEALELADKPEGY
LYFTIPNPTKYNLLGAGVGIGSTHFRVEVEAFYDKFNLTAPAGYLHNKFYEYFA
LATMDTHQPSAEDKYIWMKNTGITISPFILNACDYFILKTRNVPAPICLQVGGN
FIDLDQFHPQAOAKYGIYFVSPNIAFFIDGSHGLNMQFSDSPVVDYSSGFP
TISAKFANFLTSSIGRIS"
4524. .5381
/gene="p28-3"
4524. .5381
/gene="p28-3"
/translation="component of 28-kDa outer membrane protein
multigene family"
/codon_start=1
/transl_table=11
/product="p28-3"
/protein_id="AAF73411.1"
/db_xref="GI:8132828"
/translation="MQLYISFIILSGLLPKYVFCMHQNNIDGSYVTIKYQLTTPH
FNKYIKETDFTQPIGLAKITANTKDTLKENFSPFLHOTSQYKSYQNDLLIGL
SVGLPVKSFRIEFEGAKYNTKRLARYKSDGYKYFAI PRKSEHGFLDNTFGYTVAK
NNGIISIINILCSETKYKFTPYICIGVGDDFIEIDVMRIKPAYQGVGVSPIT
SKLIUSGQYHKVIGNKRFELLPVYQPVKELKLVNTKYSKOIDQDVTASLTINLHPS
SEIGUSFI"
5394. .6212
/gene="p28-4"
5394. .6212
/gene="p28-4"
/translation="component of 28-kDa outer membrane protein
multigene family"
/codon_start=1
/transl_table=11
/product="p28-4"
/protein_id="AAF73412.1"
/db_xref="GI:8132829"
/translation="MYMNNKHVCYIVYVITLFFLLPIESLSALIGNVEKDLKVS
TYVSSQYKPSIFHFRNFIQESHKPKSEEFKKIKANLNNILKSNAYNLQDNTSF
SGTIGYFGKGLRLAEAGCYQEPNKNNSNLIISNNKYSRIHDENYAITNNKLSIA
STMVNTCYDISNNTSIVPCLCTGIGEDLVGLFNTIHFKLAYQGVGMSYILNNILL
PSDIYHKVMGNRKNLYMQYVADPNISEETIPIILAKLDIGYFGSGEIGRPFN"
6228. .7115
/gene="p28-5"
6228. .7115
/gene="p28-5"
/translation="component of 28-kDa outer membrane protein
multigene family"
/codon_start=1
/transl_table=11
/product="p28-5"
/protein_id="AAF73413.1"
/db_xref="GI:8132830"

```

Alignment Scores:  
Pred. No.:  
Score:

```

/translation="MTKKFNFNVLITLLFLPLKSFSTYANNNTITQKGLYISGQ
YKSLPHFKNFSVEENDKVLDLGLTDTVTYTEHLRDNTKNTHTYIAKFNKFIIE
SIAIGYSGGQPLEIESSGDFDVNNYKNAVQDVNRYFALVREKNGSFPSPKPHET
SQPSDNPKKSFYTLMKNGVGVVASVIINGCYDFSNNNTTISPVVICVGVDLFEFE
VMHIFACQSGVIGISPISTIFADAHYHKVNNKFNKLNHLVKYSYELKNSPTTISA
TAKLNIYFGGEGVMRFIF"
7137. .7976
/gene="p28-6"
7137. .7976
/gene="p28-6"
/translation="component of 28-kDa outer membrane protein
multigene family"
/codon_start=1
/transl_table=11
/product="p28-6"
/protein_id="AAF73414.1"
/db_xref="GI:8132831"
/translation="MSKKFTITGTVLASLLSFLSIESFSAINHNHNTGNTSGIYTG
QYRPGVSHFSNFVKETNVDITQLVGKYSASSIDPNTYSNFOQYTVTFQDAASTF
KAIGSYSPESIRLEGSYEFVKDPKDYSAKDAEFALARNTSTVTFQDAKYTM
KNGLSVASIMINGCYDFSNNLVVSPYICAGIGEDFIEFDTLHLKLAQCKLGISY
YFFPKINVPAGGYHVRVIGNKFNKLNLVNVTDPDFPKATSAVATLNVAYFGGAGVK
FTF"
8214. .8591
/translation="Orf378"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF73415.1"
/db_xref="GI:8132832"
/translation="MNCYLKGPLNLQIFYGIFSFIRNFQNTLIIINDSKCGPYTTL
WDRPALHYTTLTGSEYRNFEDILYENIICQCKLLINYSVLNQHKNKTLVIPIPN
ARFNSNIRVRNISINKESYEC"
8581. .9432
/gene="p28-7"
8581. .9432
/gene="p28-7"
/translation="component of 28-kDa outer membrane protein
multigene family"
/codon_start=1
/transl_table=11
/product="p28-7"
/protein_id="AAF73416.1"
/db_xref="GI:8132833"
/translation="MSAKKFLFIGSVLGVLPKTSLSNLNNNTKCTGLYVS
QYKPTVSHFSNFKLKYTYDTKELLGLAKDITDITTNKFNIPINTKFDNAVS
FSAVGYISQDSRVEVEYSEFVKNPGNVYVSEAFYALARGIDNLOKYPETNK
YVVIKNGLSVASIIINGCYDFSNNLVKSPYICVGFQGDIIIEFFSAVSKFAYQGVK
GISYPLFSNMIIIPADGYHVKVIGNKFNKLNLVNVTDPDFPKATSAVATLNVAYFGSE
FGLKFI"
9441. .10268
/gene="p28-8"
9441. .10268
/gene="p28-8"
/translation="component of 28-kDa outer membrane protein
multigene family"
/codon_start=1
/transl_table=11
/product="p28-8"
/protein_id="AAF73417.1"
/db_xref="GI:8132834"
/translation="MSKKNFITIGATLIHMLPNI SPETNNNTDKLSGLYISGQYK
PGISHFSKFSVKEIYNDNIQLIGLRHNAISTSLNTDNIPIKYVTFQNNITSPSGA
IGYSDPTGARPELEGSYEEDVDTPGDLCLNTDITRYFALARNKSGSSFTSNNTYMRN
DGVSIITVINGCYDIIFLKDLEVPYVGVGDDFIEFFDALHKLAYQCKLGINHL
STQASVFDIGYHVKVIGNQFNKLNLVNVTDPDFGVAVATLNTIGYFGGIGRLTF"
10278. .11135
/gene="p28-9"
10278. .11135

```

2.77e-77 Length: 14759  
1033.00 Matches: 201

Percent Similarity: 80.20% Conservatives: 34  
Best Local Similarity: 68.60% Mismatches: 56  
Query Match: 69.05% Indels: 2  
DB: 1 Gaps: 2

US-10-062-624-40 (1-293) x AF230642 (1-14759)

Qy 1 MetAsnAsnLysLeuLysPheThrIleLeuAsnThrValLeuValCysLeuLeuSerLeu 20  
Db 11160 ATGAATAAAAAACAAGTTT--ATTATAGCTACAGCATGGTATATTTACTGTCATTA 11216  
Qy 21 ProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40  
Db 11217 CCTAGTATCGTTTTCAGAGTTACAAACGAGTATTAATAAACACTCTGGGTTATAT 11276  
Qy 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60  
Db 11277 ATTAGTGACAATAACAACCAAGTGTCTGTTTTAGTAGTTCTCAATTAAAGAACT 11336  
Qy 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80  
Db 11337 AACACTATCAAAAATTTCTATAGCGTTAAAAAAGATATTAACTCTCTGAAGTTAAC 11396  
Qy 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaVal 100  
Db 11397 GCCGATGCTAGTCAAGGTATTAGTATCCAGGAATTTTACTATACCTTATATAGCACA 11456  
Qy 101 PheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThr 120  
Db 11457 TTTCAGATAATGCTTTTATTTCAACGGTGCTATTGGTTAC--ATTACTGAAGGCTA 11513  
Qy 121 ArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThr 140  
Db 11514 AGGATTGAATAGAAAGGTTCCTATGAAGATTTTCATGCTAAAAACCCCTGGAGGTATG 11573  
Qy 141 LeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThr 160  
Db 11574 CTAATGATGCTTTCGGTACTTTGCTTTAGCAGTGTATGGAAGCAACAACTGCCAA 11633  
Qy 161 ProlysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSer 180  
Db 11634 CCNAAACACAAAGCTCACAAAAGATTTTACACTGTATGAGAGTGATGGTTATCT 11693  
Qy 181 IleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnLeuSerIleSer 200  
Db 11694 ATAATATCTATCATGGGTAAACGGCTGCTATGATTTTCTTCGGATAATTTATTAGTATCA 11753  
Qy 201 ProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIle 220  
Db 11754 CCTATATATGTGGAGGTATAGGGGTGGATGCAATGAATTTTGTGACGATTACACATT 11813  
Qy 221 LysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeu 240  
Db 11814 AAACCTGGTGCCCAACAAATAGGATAGGATCACTATCAATTTATTTATATATCAGTTA 11873  
Qy 241 PheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGln 260  
Db 11874 TTTGCTGTGGATTTTATCATCAAGTAATAGTAAGTAACCAATTCAGAAATTTAAACGTTCAA 11933  
Qy 261 HisValAlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIle 280  
Db 11934 CATGTAGTGAACCTTAATGATGACCTTAAAGTTATCATCTGCAGTTGCCACACTTAATGTT 11993  
Qy 281 GlyTyrPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293  
Db 11994 GGATATTCGGCGCTGAAGTTGGAGTGAAGATTATATTT 12032

RESULT 5  
AF324792  
LOCUS  
DEFINITION  
Ehrlichia canis phosphoribosylaminoimidazole carboxylase (purK)  
gene, complete cds; major outer membrane protein gene cluster 2,  
complete sequence; and u6 gene, partial cds.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
gene  
CDS  
gene  
CDS  
gene  
CDS  
gene  
CDS

AF324792  
AF324792.1 GI:13591681  
Ehrlichia canis.  
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
Anaplasmataceae; Ehrlichia.  
1 (bases 1 to 6913)  
Cloning and characterization of multigenes encoding the  
immunodominant 30-kilodalton major outer membrane proteins of  
Ehrlichia canis and application of the recombinant protein for  
serodiagnosis  
J. Clin. Microbiol. 36 (9), 2671-2680 (1998)  
98371112  
9705412  
2 (bases 1 to 6913)  
Chashi, N., Rikihisa, Y. and Unver, A.  
Analysis of transcriptionally active gene clusters of major outer  
membrane protein multigene family in Ehrlichia canis and E.  
chaffeensis  
Infect. Immun. 69 (4), 2083-2091 (2001)  
21153566  
11254561  
3 (bases 1 to 6913)  
Chashi, N., Rikihisa, Y. and Unver, A.  
Direct Submission  
Submitted (29-NOV-2000) Veterinary Biosciences, Ohio State  
University, 1925 Coffey Rd., Columbus, OH 43210, USA  
Location/Qualifiers  
1. .6913  
/organism="Ehrlichia canis"  
/strain="Oklahoma"  
/db\_xref="taxon:944"  
329. .1420  
/gene="purK"  
329. .1420  
/genes="purK"  
/codon\_start=1  
/transl\_table=11  
/product="phosphoribosylaminoimidazole carboxylase"  
/translation="MLHNSYVSPGSTIGIGGQGMISIAATNLGYKTHLLTDNQD  
SPSINTNNVTLNDYNNKELLLEFASVDIATLEPNIPTSTIDILSOXTKVPKGT  
ALHISNRIKQYIRNLGKLTANFRIIDYNDNLVNLVLELGYPAALLKTELGVGDKG  
OYLITQGLNTLSTLWQOEYILEKFKHKEISIIISKSINGSIEFFPVAENHHTN  
GILATSPASISSEINQAKNIALQIAESFNVLGILAVEFFITDSQELIVNETAPRN  
HNSGWSLACNISOFEUTRAICGLPKPKLLFPCTWMDNLDGDDIYNYQYKTKIN  
ESTIYIKTKASKNRKMGHINRLNKHEI"  
1756. .2598  
/genes="p30-10"  
1756. .2598  
/genes="p30-10"  
/notes="P30 family"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein p30-10"  
/protein\_id="AAK31312.1"  
/db\_xref="GI:13591683"  
/translation="MNYKILVRSALISMSILPYQSFADPVGSRNDNKEGFIYSK  
YNSISGHFRKFAEETPTINGNSLTKKVFKDGDITKKDFTVPAFGIDFQNNLIS  
YPSGISGYMDGPRIELEAYQOFPNKNTDNDTNGEYKHFALSRKDAEDQOYV  
LKNDGITFMSLWNTCYDITAEQVSFVPYACAGIADLITFKDLNLPAYQKIGIS  
YPIPEVSATFISGYHGVIGNKEPIVITPVVINDAPQTTTSVTLVDVGYFGEIGM  
RTTF"  
2926. .3756  
/genes="p30-4"  
2926. .3756  
/genes="p30-4"  
/note="P30 family"  
/codon\_start=1  
/transl\_table=11



```

/organism="Ehrlichia ruminantium"
/isolate="Highway"
/db_xref="taxon:779"
948..1796
/notes="ORF2"
/codon_start=1
/transl_table=11
/product="major antigenic protein 1 like protein"
/protein_id="AAD26347.1"
/db_xref="GI:4589101"
/transl_table=11
/transl_table=11
SAKYNPSIPIHFRKFSAEETPVYVKDSTPKVGLKKEGSLTKYSDPTDTSREGONN
FISGSSIGYVMDPRVEIAAYQKFNPNANETDTSYHYKHGLSRAETWTDKY
VLTNNQVTSLSMNFNACYDITAEGVFFIPACAGIGADLISIPDDINLRFAYOGKIG
ISVPTPEISAFITGGYHYGVGNKYNKVPVKLPVLTLDAPQSTASVTLDDAGYFGGSL
GVRTFF"
2155..3018
/gene="MAP1"
2155..3018
/gene="MAP1"
/codon_start=1
/transl_table=11
/product="major antigenic protein 1"
/protein_id="AAD26348.1"
/db_xref="GI:4589102"
/transl_table=11
MPTASHFGKMSIKEDSKNTQTVFGLKQMDGVKTPSDSNTNSTIFTKQSYFREN
PFLGFAIGYVSNPGYETPEVSEYVETDVKRNPNGYKNNAHMYCALDTAAQNSNGAG
LTSVMVKNLNTISILMNAICYDILMDGIPVSPYVCAGIGTDLVSINATNPKLSYQ
GKLGISVINSSEASIFIGGHPHRVIGNEFKDIATLKIFTSKTSIGNSPGFASALDVCH
FGIEIGGRFVP"
BASE COUNT 1288 a 473 c 459 g 1287 t
ORIGIN

Alignment Scores:
Pred. No.: 1,27e-31 Length: 3507
Score: 484.50 Matches: 114
Percent Similarity: 53.87% Conservative: 53
Best Local Similarity: 36.77% Mismatches: 100
Query Match: 32.39% Indels: 43
DB: 1 Gaps: 11

US-10-062-624-40 (1-293) x AF125276 (1-3507)

Qy 1 MetAsnLysLeuLysPheThrIleAsnThrValLeuValCysLeuLeuSer--- 19
Db 942 ATTAATATGAATTATAAAAAATTCGTGAAGAGTGGTTAATCTCATTAACTCATTT 1001
Qy 20 LeuPro-----AsnIleSerSerLysLysAlaIleAsnAsnAla 33
Db 1002 TTACCATATCAGTCTTTTGCAGAACCTGTAAGTTCAAAATACATTTGGAATGAAATGCT 1061
Qy 34 LysLysTyrTyrGlyLeuTyrIleSerGlyGlnTyrLysProSerValSerValPheSer 53
Db 1062 AAAGAA-----GGTTCTACATGAAGTGCAAAATACCAACCAAGCATACCACACTTCAGA 1115
Qy 54 AsnPheSerValLysGluThrAsnValIle-----ThrLysAsnLeuIle 68
Db 1116 AAATTTTCTGTCAGGAAGAACTCCTGTATACGGTAAGAACTCCAACTAAAGGTATTT 1175
Qy 69 AlaLeuLysLysAspValAlap-----SerIleGluThrLysThrAspAlaSer 84
Db 1176 GGGTTAAAGAAAGGAGGGTCTTATAACAAATATACAGTGATTTCACTAGAACAGAT----- 1229
Qy 85 ValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaValPheGlnAspAsn 104
Db 1230 ATATCGTTGAGGCCCAAAATAATTTTATC----- 1259
Qy 105 SerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArgValGluIle 124
Db 1260 ---TCAGTTTCTCAGGACAGCATAGTTAT---ATCATGTGATGGACCAAGAGTAGATT 1313
Qy 125 GluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThr---LeuSerAsp 143

```

```

Db 1314 GAAGCTGCATACCAAAAAATTCACCCAAAAATCCAGCTAATGAACCTGACTAGTAGT 1373
Qy 144 AlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGlu 163
Db 1374 TACTATAACACTATGGATTATCTCGT-----GCAGAAACCATGACAGATAAAAA 1424
Qy 164 LysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleSer 183
Db 1425 -----TATGTTGTACTTACAAATTAATGGAGTAACCTTTTTCATCA 1463
Qy 184 ValIleValAsnValCysTyrAspPheSerLysLeuAsnAsnLeuSerIleSerProTyrIle 203
Db 1464 TTAATGTTAATGCTGCTATGACATTACAGCAGAGAGGAGTACCTTTTCATTCATATGCA 1523
Qy 204 CysGlyGlyAlaGlyValAspAlaIleGluPheAspValLeuHisIleLysPheAla 223
Db 1524 TGTGCTGTTATGTCGATGATCTTATATCTATATTTGATGATATAAATTTTGTCT 1583
Qy 224 TyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSer 243
Db 1584 TACCAAGGTAAATTTGGTATTAGTTTCAATTAATCTCTGAAATTTCTGCATTTATTGGC 1643
Qy 244 LeuTyrThrIleLysValMetGlyAsnGlnPheLysLeuAsnValGlnHisValAla 263
Db 1644 GGTATATTACCATGGAGTAATAGGTAACAAGTATAACAAGTACCTGTAAAGCTTCCCTGTA 1703
Qy 264 GluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIleGlyTyrPhe 283
Db 1704 ACTTTAACAGATGCTCTCTCAAGCACCTTCGCTTCAGTAACCTCTTGACGCTGGATATTT 1763
Qy 284 GlyGlyGluIleGlyAlaArgLeuThrPhe 293
Db 1764 GGTGTTGAACCTGGAGTAAGGTTTACTTTTCT 1793

RESULT 7
AF125275 3551 bp DNA linear BCT 20-APR-1999
LOCUS Cowdria ruminantium isolate Crystal Springs major antigenic protein
DEFINITION 1 like protein and major antigenic protein 1 (MAP1) genes, complete
cds.
ACCESSION AF125275
VERSION AF125275.1 GI:4589097
KEYWORDS Ehrlichia ruminantium.
SOURCE Ehrlichia ruminantium.
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE 1 (bases 1 to 3551)
AUTHORS Sulsona,C.R., Mahan,S.M. and Barbet,A.F.
TITLE The 'map1' Gene of Cowdria ruminantium Is a Member of a Multigene
Family Containing Both Conserved and Variable Genes
JOURNAL Biochem. Biophys. Res. Commun. 257 (2), 300-305 (1999)
REFERENCE 2 (bases 1 to 3551)
AUTHORS Sulsona,C.R., Mahan,S.M. and Barbet,A.F.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-1999) College of Veterinary Medicine Department
of Pathobiology, University of Florida, P.O. Box 110880,
Gainesville, FL 32610, USA
FEATURES
source Location/Qualifiers
1..3551
/organism="Ehrlichia ruminantium"
/isolate="Crystal Springs"
/db_xref="taxon:779"
991..1839
/notes="ORF2"
/codon_start=1
/transl_table=11
/product="major antigenic protein 1 like protein"
/protein_id="AAD26345.1"
/db_xref="GI:4589098"
/translation="MNYKKILVRSALISLMSFLPYQSFAFPVSNINIGNNAKEGFYI
SAKYNPSIPIHFRKFSAEETPVYVKDSTPKVGLKKEGSLTKYSDPTDTSREGONN
FISGSSIGYVMDPRVEIAAYQKFNPNANETDTSYHYKHGLSRAETWTDKY
VLTNNQVTSLSMNFNACYDITAEGVFFIPACAGIGADLISIPDDINLRFAYOGKIG
ISVPTPEISAFITGGYHYGVGNKYNKVPVKLPVLTLDAPQSTASVTLDDAGYFGGSL
GVRTFF"
CDS

```

FISFGSGIGVIMDPVRVEIAEAAOKNPKNPANETDTSYXKHVGLSRAETWTDKKY  
 WLTNNGVTFSSLMFNACDYITAGVPFPIYACAGIGADLLISIEDDDINLKPAYOCKIG  
 ISYPTTPEISAFIGGYHGVIGNKYNKVPVKLPVTLTDAPOSTASVTLDAGYFGGEL  
 GVRFTF"

Gene

2199..3062

CDS

/gene="MAP1"

2199..3062

/gene="MAP1"

/codon\_start=1

/transl\_table=11

/product="major antigenic protein 1"

/protein\_id="AAD26346.1"

/db\_xref="GI:4589099"

/translation="MNCKKIFITSTLISVFLGCVSFSDVIOEDSNPAGSVYISAKY

PFLTAFAGIKMSIKEDSKNTQVFLKKDWDGKVPKPSDNTNNTFTFKDYSPFYENN

MPTFAGAGIYSNMGPRIEFEVYETFDVKNPGNYKNNAMHYCALDIAQNSYNGAG

LTTSMVKNENLTNLSMLNACYDILMDGIPVPCVACIGITGLVSVINATNPGLSVQ

GKLGISYSINSEASIFIGGHPHRVIGNEFKDIATLKIFTSKTIISNPGFASATLDVCH

FGIEIGGRFVF"

BASE COUNT 1308 a 477 c 464 g 1302 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 1.29e-31 Length: 3551  
 Score: 484.50 Matches: 114  
 Percent Similarity: 53.87% Conservative: 53  
 Best Local Similarity: 36.77% Mismatches: 100  
 Query Match: 32.39% Indels: 43  
 DB: 1 Gaps: 11

US-10-062-624-40 (1-293) x AF125275 (1-3551)

Qy 1 MetAenAenLysLeuLysPheThrIleleAenThrValLeuValCysLeuLeuSer--- 19  
 Db 985 ATTAATATGAATATATAAAAAAATCTCGTAAGAAGTGGTTAAATCTCATTAAATGTCATT 1044  
 Qy 20 LeuPro-----AsnIleSerSerSerLysAlaIleAenAenAenAla 33  
 Db 1045 TTACATATACGTCTTTTGCAGAACCTGTAGTTCAAATAACATTTGGAATGAAATGCT 1104  
 Qy 34 LysLysTyTyTyGlyLeuTyLysSerGlyGlnTyLysProSerValSerValPheSer 53  
 Db 1105 AAAGAA-----GGGTTCTACATAGTGCANAATACACCCAGCATACACACTTCAGA 1158  
 Qy 54 AenPheSerValLysGluThrAenValIle-----ThrLysAenLeulle 68  
 Db 1159 AAATTTTCTGCTGAGGAAACTCTGTATACGGTAAAGACTCTCCAACTAAAGAGTATT 1218  
 Qy 69 AlaLeuLysLysAspValAsp-----SerIleGluThrLysThrAspAlaSer 84  
 Db 1219 GGGTTTAAAGAGGAGGGTCTTATACAAATAACAGTGTATTCCTACGTAGAACAGAT 1272  
 Qy 85 ValGlyIleSerAenProSerAenPheThrIleProTyThrAlaValPheGlnAspAen 104  
 Db 1273 ATATCGTTTGGAGGCCAAATAATTTTATC----- 1302  
 Qy 105 SerValAenPheAenGlyThrIleGlyTyThrPheAlaGluGlyThrArgValGluLe 124  
 Db 1303 ---TCAGGTTTCTCAGGAAGCATAGGTTAT---ATCATGGTGGACCAAGAGTAGAGATT 1356  
 Qy 125 GluGlySerTyTyGluLysPheAspValLysAenProGlyGlyTyThr---LeuSerAsp 143  
 Db 1357 GAAGCTGCATACCAAAATTCACCCAAAATCCAGCTAATGAACATGATACGTAGTAT 1416  
 Qy 144 AlaTyArgTyTyPheAlaLeuAlaArgGluMetLysGlyAenSerPheThrProLysGlu 163  
 Db 1417 TACTATAAACACTATGAGTATCTCGT-----GCAGAAACCCACACAGATAAAAA 1467  
 Qy 164 LysValSerAenSerIlePheHisThrValMetArgAenAspGlyLeuSerIleleSer 183  
 Db 1468 -----TATGTTGTTACTTACAAAATAATGAGTAGTAACCTTTTTCATCA 1506  
 Qy 184 VallleValAenValCysTyTyAspPheSerAenAenLeuAenLeuSerIleSerProTyIle 203

Db 1507 TTAATGTTTAAATGCTGCTATGACATTACAGCAGAGGAGTACCTTTTCATTCATATGCA 1566  
 Qy 204 CysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIleLysPheAla 223  
 Db 1567 TGTGCTGTTATGTCGAGATCTTATATCTATATTTGATGATATAAATTTAAAAATTTGCT 1626  
 Qy 224 TyGlnSerLysLeuGlyIleAlaTySerLeuProSerAenIleSerLeuPheAlaSer 243  
 Db 1627 TACCAAGGTAAATTTGGTATTAGTTATTCCTCAATTAATCTCTGCAATTTTTCGATTATTGGC 1686  
 Qy 244 LeuTyTyTyHisLysValMetGlyAenGlnPheLysAenLeuAenValGlnHisValAla 263  
 Db 1587 GGATATTACCATGAGTAATAGTAAGTAACAGTATAACAAAGTACCTGTAAAGCTTCCTGTA 1746  
 Qy 264 GluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAenIleGlyTyPhe 283  
 Db 1747 ACTTTAACAGATGCTCTCTCAAGCACCTCCGCTTCAGTAACCTCTTGACGCTGGATATTT 1806  
 Qy 284 GlyGlyGluLeuGlyAlaArgLeuThrPhe 293  
 Db 1807 GGTGCTGAACCTGGAGTAAGGTTTACTTTC 1836

RESULT 8  
 CRUS0834  
 LOCUS 1101 bp DNA linear BCT 14-JUL-1996  
 DEFINITION Cowdria ruminantium Nyatsanga major antigenic protein 1 (map1)  
 gene, complete cds.  
 ACCESSION US0834  
 VERSION US0834.1 GI:1418257  
 KEYWORDS heartwater rickettsia strain=Nyatsanga.  
 SOURCE Ehrlichia ruminantium  
 ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 Anaplasmataceae; Ehrlichia.  
 REFERENCE 1 (bases 1 to 1101)  
 AUTHORS Reddy,G.R., Sulsona,C.R., Harrison,R.H., Mahan,S.M., Burridge,M.J.  
 and Barbet,A.F.  
 TITLE Sequence heterogeneity of the major antigenic protein 1 genes from  
 Cowdria ruminantium isolates from different geographical areas  
 Clin. Diagn. Lab. Immunol. 3 (4), 417-422 (1996)  
 JOURNAL 96400830  
 MEDLINE 8807206  
 PUBMED 2 (bases 1 to 1101)  
 REFERENCE Reddy,G.R.  
 AUTHORS Direct Submission  
 TITLE Submitted (06-MAR-1996) Roman G. Reddy, Pathobiology, University of  
 Florida, 471 Mowry Rd., Gainesville, FL 32611-0880, USA  
 JOURNAL Location/Qualifiers  
 FEATURES source  
 1..1101  
 /organism="Ehrlichia ruminantium"  
 /strain="Nyatsanga"  
 /db\_xref="taxon:779"  
 208..1080  
 /gene="map1"  
 208..1080  
 /gene="map1"  
 /notes="surface protein"  
 /codon\_start=1  
 /transl\_table=11  
 /product="major antigenic protein 1"  
 /protein\_id="AAC44145.1"  
 /db\_xref="GI:1418258"  
 /translation="MNCKKIFITSTLISVFLGCVSFSDVIOEDSNPAGSVYISAKY  
 MPTFAGAGIKMSIKEDSKNTQVFLKKDWDGKVPKPSDNTNNTFTFKDYSPFYENN  
 PFLTAFAGIYSNMGPRIEFEVYETFDVKNPGNYKNNAMHYCALDIAQNSYNGAG  
 LTTSMVKNENLTNLSMLNACYDILMDGIPVPCVACIGITGLVSVINATNPGLSVQ  
 GKLGISYSINSEASIFIGGHPHRVIGNEFKDIATLKIFTSKTIISNPGFASATLDVCH  
 VCHFGIEIGGRFVF"

BASE COUNT 379 a 159 c 170 g 394 t  
 ORIGIN

Alignment Scores:



```
Pred. No.: 8,62e-32 Length: 1101
Score: 479.50 Matches: 109
Percent Similarity: 56.11% Conservative: 61
Best Local Similarity: 35.97% Mismatches: 110
Query Match: 32.05% Indels: 23
DB: 1 Gaps: 8

US-10-062-624-40 (1-293) x CRU50834 (1-1101)

Qy 2 AsnAsnLysLeuLysPheThrIleAseThrValLeuValCysLeuLeuSer---Leu 20
Db 205 AATATGAATTCGAAGAAATTTTATCACAAGTACACTAATATCATAGTGTCAATTTTA 264
Qy 21 ProAsnLysSerSerLysAlaIleAsnAsnAsnAlaLysLysTyTyGlyLeuTy 40
Db 265 CCTGGTGTGCTCTTCTGATGTAATACAGGAAGACAGCAGCCAGCAGGAGTGTAC 324
Qy 41 IleSerGlyGlnTyLysProSerValSerValPheSerAsnPheSerValLysGluThr 60
Db 325 ATTAGCGCAATATCATGCCAATGCGATCATATTTGGTAAATGTCAATCAAAAGAGAT 384
Qy 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80
Db 385 TCAGAAATATCTCAACAGTATTTGGTCTAAAGAAAGATTGGGATGGCTTAAGTACCA 444
Qy 81 ThrAspLysValGlyIleSerAsnProSerAsnPheThrIle-----Pro 96
Db 445 ACATCA-----GAAACAGTAGTAACAATTTCTACAAATTTTACTGAAAAAGAT 489
Qy 97 TyThrAlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyThr 115
Db 490 TATTCCTTCAGATATGAAACAGTATTCGGTTTGTGTCGACGAATTTGGATCTCA 549
Qy 116 PheAlaGluGlyThrArgValGluIleGluGlySerTyGluGluPheAspValLysAsn 135
Db 550 ATG---AATGGCCAGAAATAGATTTGAAGTATCTTATGAACTTTTGTGTCAAAAT 606
Qy 136 ProGlyGlyTyThrLeuSerAspAlaTyArgTyThrPheAlaLeuAlaArgLysMetLys 155
Db 607 CCAGTGCGCACTATAAAATGATGCACACATGCTACTGTGCTTTA-----651
Qy 156 GlyAsnSerPheThrProLysGluLysValSerAsnSerIlePheHisThrValMet--- 174
Db 652 ---GATACACAGCAGCAGCACTAGTAATCAAGGTGCAACATTAGCTTCATCTGTATGTA 708
Qy 175 ArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCysTyAspPheSerLeu 194
Db 709 AAAATGAAATTTAACAGATATATCATTAATGTAATGCAATGTTATGATATATATGCTT 768
Qy 195 AsnAsnLeuSerIleSerProTyIleCysGlyGlyAlaGlyValAspAlaIleGluPhe 214
Db 769 GATGGAATGCCAGTTTCTCCATATGATGTGCAGGTATTGCTAGTTAGTGTGAGTA 828
Qy 215 PheAspValLeuHisIleLysPheAlaTyGlnSerLysLeuGlyIleAlaTySerLeu 234
Db 829 ATTAATGCTCAAAATCCTAAATATTCATATCAAGCAAGCTAGGCATTAAGTACTCAAT 888
Qy 235 ProSerAsnIleSerLeuPheAlaSerLeuTyTyHisLysValMetGlyAsnGlnPhe 254
Db 889 AATCCTGAGCTTCTATCTTTATTTGGTGGACATTTCCATAGAGTATAGGTAATGAATTT 948
Qy 255 LysAsnLeuAsnValGlnHisVal-----AlaGluLeuAlaSerIleProLys 270
Db 949 AAAGATATTGCTACCTCAAAAATTTTACTTCAACTAGTAAGCATCATCTATACCTAAT 1008
Qy 271 IleThrSerAlaValAlaThrLeuAsnIleGlyTyThrPheGlyGlyGluIleGlyAlaArg 290
Db 1009 CCTGGCTTGCATCAGCAACACTTGATGTCTGCCATTTCCGTATAGAAATTTGGAGGAGG 1068
Qy 291 LeuThrPhe 293
Db 1069 TTTGTAATTT 1077
```

RESULT 9  
AF125274  
LOCUS  
DEFINITION  
COWDRIA ruminantium isolate Welgevonden major antigenic protein 1 like protein and major antigenic protein 1 (MAP1) genes, complete cds.  
AF125274 GI:4589094  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Ehrlichia ruminantium.  
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Anaplasmataceae; Ehrlichia.  
REFERENCE  
AUTHORS  
TITLE  
The map1 Gene of Cowdria ruminantium Is a Member of a Multigene Family Containing Both Conserved and Variable Genes  
Biochem. Biophys. Res. Commun. 257 (2), 300-305 (1999)  
REFERENCE  
AUTHORS  
TITLE  
Direct Submission  
Submitted (02-FEB-1999) College of Veterinary Medicine Department of Pathobiology, University of Florida, P.O. Box 110880, Gainesville, FL 32610, USA  
JOURNAL  
FEATURES  
source  
Location/Qualifiers  
1. 3535  
/organism="Ehrlichia ruminantium"  
/isolate="Welgevonden"  
/db\_xref="taxon:779"  
987.1835  
/notes="ORF2"  
/codon\_start=1  
/transl\_table=11  
/product="major antigenic protein 1 like protein"  
/protein\_id="AAD26343.1"  
/db\_xref="GI:4589095"  
/translation="MNYKILVRSALISMSFLPYOFAEPVSSNNIGNAKVGFYI  
SAKYNPSIPIHFRKFSABETPVYKQSDPTKPKVGLKKEGSIKTSDFTRTDSIFGQNN  
FISFGSGSIGYIMDGPRVIEAAYQKFPNKPANETDTSYKHYGLSRANADKKY  
VLTNNGVTFFSLMFNACVDITAEGVFPFIPYACAGIGADLISIDFDLNLKAPYOGKIG  
ISVPTIPEISAFIGGYHGVGNKYNKVPVKLPVTLTDAPOQSTASVTLDAFYGGEL  
GVRFPT"  
2210.3055  
/genes="MAP1"  
2210.3055  
/genes="MAP1"  
/codon\_start=1  
/transl\_table=11  
/product="major antigenic protein 1"  
/protein\_id="AAD26344.1"  
/db\_xref="GI:4589096"  
/translation="MNCRKIFITPILSLVPLGVSPDAIOEDSSPVGSVVISAKY  
MPTASHFGKMSIKEDSRDTKVVEGLKDWGVKTSSTNTIFTKDYKSFYKNNPFLGF  
AGAIGYFMGPRIEFISYETFDVKNPKNPKNDAMHYCALDTATSSGGAASVNV  
KINLEPDI SLMLNACNADIMLDGPNVPYVYAGIGTDLVSVINSPKLSYOGKIGISY  
SNPEASITIGGHFHRVIGNEFKDITTSKIENSTGATPGGPASATLDVCHFGIEIG  
GRFVF"  
BASE COUNT 1290 a 445 c 480 g 1320 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3,41e-31 Length: 3535  
Score: 479.50 Matches: 113  
Percent Similarity: 53.55% Conservative: 53  
Best Local Similarity: 36.45% Mismatches: 101  
Query Match: 32.05% Indels: 43  
DB: 1 Gaps: 11  
US-10-062-624-40 (1-293) x AF125274 (1-3535)  
Qy 1 MetAsnAsnLysLeuLysPheThrIleAseThrValLeuValCysLeuLeuSer--- 19  
Db 981 ATTAATGAATTAATAAAAAATTCGTAAGAAGTCGCTTAATCTCGTTAATGTCATTT 1040



```

Qy 69 AlaLeuLysLysAspValasp-----SerlleGluThrLysThrAspAlaSer 84
Db 1219 GGGTTAAAGAGGAGGGTCTATACAAATAACAGTGAATTCACAGTAAACAGAT----- 1272
Qy 85 ValGlylleSerAsnProSerAsnPheThrIleProTyrThrAlaValPheGlnAspAsn 104
Db 1273 ATATCGTTGAGGGCCAAATAATTTTATC----- 1302
Qy 105 SerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArgValGluIle 124
Db 1303 ---TCAGGTCTTCAGGAAGCATAGGTAT---ATCATGGATGGACCAAGAGTAGAGATT 1356
Qy 125 GluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThr---LeuSerAsp 143
Db 1357 GAAGCTGCATACCAAAATTCACCAAAATCCAGCTAATGAACATGATAGTATGAT 1416
Qy 144 AlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGlu 163
Db 1417 TACTATAACACATAGGATTAATCTCGT-----GCAGAGCTATGCGAGATNAAAA 1467
Qy 164 LysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSer 183
Db 1468 -----TATGTTGCTACTTACAAATAATGGAGTAACATTTTTCATCA 1506
Qy 184 VallleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSerProTyrIle 203
Db 1507 TTAATGTTTCAATGCTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1566
Qy 204 CysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIleIlePheAla 223
Db 1567 TGTGCTGATGTTGGTGGAGATCTATATCTATATTTGATGATGATGATGATGATGATGATGAT 1626
Qy 224 TyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSer 243
Db 1627 TACCAAGGTAAATGGTATTAGTTATCTCAATTAATCTCTGAAATTTCTGCAATTTATGGC 1686
Qy 244 LeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGlnHisValAla 263
Db 1687 GGATATTACCATGGAGTATAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1746
Qy 264 GluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIleGlyTyrPhe 283
Db 1747 ACTTTAACAGATGCTCTCTCAAGCACTTCCGCTCAGTAATCTTTCAGCTGGATATTTT 1806
Qy 284 GlyGlyGluIleGlyAlaArgLeuThrPhe 293
Db 1807 GGTGCTGAACCTGGAGTAAGGTTTACTTTTC 1836

RESULT 11
AF125279 3541 bp DNA linear BCT 20-APR-1999
LOCUS Cowdria ruminantium isolate Antigua major antigenic protein 1 like
DEFINITION protein and major antigenic protein 1 (NAP1) genes, complete cds.
ACCESSION AF125279
VERSION AF125279.1 GI:4589109
KEYWORDS
SOURCE Ehrlichia ruminantium.
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE
1 (bases 1 to 3541)
AUTHORS Sulsona,C.R., Mahan,S.M. and Barbet,A.F.
TITLE The map1 Gene of Cowdria ruminantium Is a Member of a Multigene
Family Containing Both Conserved and Variable Genes
JOURNAL Biochem. Biophys. Res. Commun. 257 (2), 300-305 (1999)
REFERENCE
2 (bases 1 to 3541)
AUTHORS Sulsona,C.R., Mahan,S.M. and Barbet,A.F.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-1999) College of Veterinary Medicine Department
of Pathobiology, University of Florida, P.O. Box 110880,
Gainesville, FL 32610, USA
FEATURES
Location/Qualifiers
source 1..3541

```

```

/organism="Ehrlichia ruminantium"
/isolate="Antigua"
/db_xref="taxon:779"
989..1837
/notes="ORF2"
/codon_start=1
/transl_table=11
/product="major antigenic protein 1 like protein"
/protein_id="AAD26353.1"
/db_xref="GI:4589110"
/transl_table="MNYKKILRSALISLMSFLPQSPFAEPVSSNNIGNENAKGFIY
SAKNYPIPHRPFKESASSETPVYKDSPTKKVFGGLKXSGSITKYSDFTRTDTSFQGN
FISGSGISGIMDGPVEIEAAATOKENPNPANETDSDIYKHGSLSRAMADKKY
VLSNNGVTTFSSLMFNACYDITAEGVPIFYACAGIGADLISIFDDINLRFAYQKIG
ISYPTTPEISAFIGGYTHGVGNKYNKVPVKLPVLTLDAPQSTASVTLDAFYGGEL
GVRFTF"
2210..3055
/gene="NAP1"
2210..3055
/gene="NAP1"
/transl_table=11
/product="major antigenic protein 1"
/protein_id="AAD26354.1"
/db_xref="GI:4589111"
/transl_table=11
/product="major antigenic protein 1"
/protein_id="AAD26354.1"
/db_xref="GI:4589111"

BASE COUNT 1282 a 451 c 479 g 1329 t
ORIGIN
Alignment Scores:
Pred. No.: 3,41e-31 Length: 3541
Score: 479.50 Matches: 113
Percent Similarity: 53.55% Conservative: 53
Best Local Similarity: 36.45% Mismatches: 101
Query Match: 32.05% Indels: 43
DB: 1 Gaps: 11

US-10-062-624-40 (1-293) x AF125279 (1-3541)
Qy 1 MetAsnAsnLysLeuLysPheThrIleIleAsnThrValLeuValCysLeuLeuSer--- 19
Db 983 ATTAATATGAATTATAAAAAATTCGTGAGAAGTGGCTTAATCTCATTATGTCATT 1042
Qy 20 LeuPro-----AsnIleSerSerLysAlaIleAsnAsnAsnAla 33
Db 1043 TTACCATATCAGTCTTTTGCAGAACCTGTAAGTTCAAAATCAATTTGGAATGAAATGCT 1102
Qy 34 LysLysTyrTyrGlyLeuTyrIleSerGlyClnTyrLysProSerValSerValPheSer 53
Db 1103 AAAGAA-----GGGTTCTACATAAGTGCAAAATACAAACCCAGCATACCACTTCA 1156
Qy 54 AsnPheSerValLysGluThrAsnValIle-----ThrLysAsnLeuIle 68
Db 1157 AAATTTTCTGCTGAGGAACCTCTGTATACGGTAAAGACTCTCCAACTAAAGGTATTT 1216
Qy 69 AlaLeuLysLysAspValasp-----SerlleGluThrLysThrAspAlaSer 84
Db 1217 GGGTTAAAGAGGAGGGTCTTATAACAAATAACAGTGAATTCACAGTAAACAGAT----- 1270
Qy 85 ValGlylleSerAsnProSerAsnPheThrIleProTyrThrAlaValPheGlnAspAsn 104
Db 1271 ATATCGTTGAGGGCCAAATAATTTTATC----- 1300
Qy 105 SerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArgValGluIle 124
Db 1301 ---TCAGGTCTTCAGGAAGCATAGGTAT---ATCATGGATGGACCAAGAGTAGAGATT 1354
Qy 125 GluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThr---LeuSerAsp 143

```

```

1355 GAAGCTGCATCAAAATTCACCCAAAATCCAGCTAATGAACCTGATAGTATGAT 1414
144 AlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGlu 163
1415 TACTATAAACACTATGATTATCTCGT-----GCAGAAGCTATGCGAGATAAAAA 1465
164 LysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleSer 183
1466 -----TATGTTGTACTTACAAATAATGAGTAACCTTTTCATCA 1504
164 ValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSerProTyrIle 203
1505 TTAATGTTCAATGCTTATGATGATTAACGAGAGAGTACCTTTTCATCCATGCA 1564
204 CysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIleLysPheAla 223
1565 TGTGCTGGTATTGGTGCAGACTTATATCTATATTGATGATATAAATTTAAATTCGCT 1624
224 TyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSer 243
1625 TACCAAGGTAAATGGTATTAGTTATCCAAATTTACTCTCGTGAATTTCTGCAATTTATGCGC 1684
244 LeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGlnHisValAla 263
1685 GGATATTACCTGGAGTATAGTAAAGTATACAAAGTATACAAAGTACCTGTAAGCTTCTGTA 1744
264 GluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIleGlyTyrPhe 283
1745 ACTTTAACAGATGCTCTCTCAAGCACTTCGGCTTCAGTAACCTTTCGACGCTGATATTTT 1804
284 GlyGlyGluIleGlyAlaArgLeuThrPhe 293
1805 GGTGGTGAACCTGGAGTAAGTTTACTTTC 1834

RESULT 12
AF125278 3572 bp DNA linear BCT 20-APR-1999
LOCUS Cowdria ruminantium isolate Umbanein major antigenic protein 1 like
DEFINITION protein and major antigenic protein 1 (MAP1) genes, complete cds.
ACCESSION AF125278
VERSION AF125278.1 GI:4589106
KEYWORDS
SOURCE Ehrlichia ruminantium.
ORGANISM Ehrlichia ruminantium
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE
1 (bases 1 to 3572)
Sulsona, C.R., Mahan, S.M. and Barbet, A.F.
The map1 Gene of Cowdria ruminantium Is a Member of a Multigene
Family Containing Both Conserved and Variable Genes
JOURNAL Biochem. Biophys. Res. Commun. 257 (2), 300-305 (1999)
AUTHORS Sulsona, C.R., Mahan, S.M. and Barbet, A.F.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-1999) College of Veterinary Medicine Department
of Pathobiology, University of Florida, P.O. Box 110880,
Gainesville, FL 32610, USA
FEATURES
Location/Qualifiers
source
1. 3572
/organism="Ehrlichia ruminantium"
/isolates="Umbanein"
/db_xref="taxon:779"
989..1837
/note="ORF2"
/codon_start=1
/translation="major antigenic protein 1 like protein"
/product="major antigenic protein 1 like protein"
/protein_id="AAD26351.1"
/db_xref="GI:4589107"
translation="MNYKKILVRSALISLMSFLPYQSFAPVSSNNIGNENAKEGFYI
SKAYNPSIPIHFRKFSAEPTPVYKDSPTKKVFLGKKEGSIKYSDFTRIDISFEGNN
FISGFSGISGYIMDGRVREIAAYQKFNPNANETDTSYYKYHGLSRAEMADKKY

```

```

VLTNGVTSSLMFNACYDITAGVPPFIPVACAGIGADLISIFDDINLKPAYQKIG
ISYPTIETSAFIGGYHGVGNKYNKVPVKLPVTLTDAQSTASVTLDAQYFGEGL
GVRFTF"
2210..3082
/gene="MAP1"
2210..3082
/gene="MAP1"
/codon_start=1
/translation="major antigenic protein 1"
/product="major antigenic protein 1"
/protein_id="AAD26352.1"
/db_xref="GI:4589108"
/translation="MNCCKIFITSLISLVSPGVPSFDVIOESNPNAGSVYISAKY
MPTASHFGKMSIKEDSKNTQVFLGKDMGVKTPSSDSGNSIIFTEKDYSPYENN
PFLGAGAGISYMNPGPRIEFVETFDVKNPGNYKNDAMHYCALDTGTGTSQTGAT
LNSVMKNENLTDIALMNACYDITILEGMPVSPYCVAGIGTDILVSNINATNPKLSYQ
KLGISYINPEASIFIGGHFHRVIGNEFKDITTSKIPSTGKLATAASPQASATLD
VCHFGIIGGRFVF"
BASE COUNT 1303 a 472 c 476 g 1321 t
ORIGIN
Alignment Scores:
Pred. No.: 3,45e-31 Length: 3572
Score: 479.50 Matches: 113
Percent Similarity: 53.55% Conservative: 53
Best Local Similarity: 36.45% Mismatches: 101
Query Match: 32.05% Indels: 43
DB: 1 Gaps: 11
US-10-062-624-40 (1-293) x AF125278 (1-3572)
QY 1 MetAsnAsnLysLeuLysPheThrIleIleAsnThrValLeuValCysLeuLeuSer--- 19
983 ATTAATATGAATTAATAAAAAATTCGTGAAGAAGTGGTAAATCTCAATTAATGTCATT 1042
QY 20 LeuPro-----AsnIleSerSerLysAlaIleAsnAsnAla 33
1043 TTACCATATCAGTCTTTTGCAGAACCTGTAGTTCAATAACATTCGGAATCAAAATGCT 1102
QY 34 LysLysTyrTyrGlyLeuTyrIleSerGlyGlnTyrLysProSerValSerValPheSer 53
1103 AAAGAA-----GGGTTCTACATAAGTCGAAATAACACCAACCAACCACTCCAGATTCAGA 1156
QY 54 AsnPheSerValLysGluThrAsnValIle-----ThrLysAsnLeuIle 68
1157 AAATTTTCTGCTGAGGAACCTCTGTATACGGTAAAGACTCTCCAACTAAAGAGTATTT 1216
QY 69 AlaLeuLysLysAspValAsp-----SerIleGluThrLysThrAspAlaSer 84
1217 GGGTTAAAAAGGAGGGTCTATAACAAAATACAGTGTATTTCTACCTAGAACAGAC----- 1270
QY 85 ValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAlaValPheGlnAspAsn 104
1271 ATATCGTTTGGAGGGCAAAATAAATTTTATC----- 1300
QY 105 SerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArgValGluIle 124
1301 ---TCAGGTTTCTCAGGAAGCATAGGTAT---ATCATGGATGGACCAAGAGTAGAGATT 1354
QY 125 GluGlySerTyrGluGluPheAspValLysAsnProGlyGlyTyrThr---LeuSerAsp 143
1355 GNAGCTGCATACCAAAAATTCACCCAAAATAATCCAGCTAATGAACACTGACACTAGTAT 1414
QY 144 AlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGlu 163
1415 TACTATAAACACTATGATTATCTCGT-----GCAGAAGCTATGCGAGATAAAAA 1465
QY 164 LysValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleSer 183
1466 -----TATGTTGTACTTACAAATAATGAGTAACCTTTTCATCA 1504
QY 184 ValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSerProTyrIle 203

```



LOCUS CRU50835 1278 bp DNA linear BCT 14-JUL-1996  
 DEFINITION Cowdria ruminantium Um Banein major antigenic protein 1 (map1)  
 gene, complete cds.  
 ACCESSION U50835  
 VERSION U50835.1 GI:1418259  
 KEYWORDS heartwater rickettsia strain=Um Banein.  
 SOURCE Ehrlichia ruminantium  
 ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 Anaplasmataceae; Ehrlichia.  
 REFERENCE 1 (bases 1 to 1278)  
 AUTHORS Reddy, G.R., Sulsona, C.R., Harrison, R.H., Mahan, S.M., Burridge, M.J.  
 TITLE Sequence heterogeneity of the major antigenic protein 1 genes from  
 Cowdria ruminantium isolates from different geographical areas  
 JOURNAL Clin. Diagn. Lab. Immunol. 3 (4), 417-422 (1996)  
 MEDLINE 96400830  
 PUBMED 8807206  
 REFERENCE 2 (bases 1 to 1278)  
 AUTHORS Reddy, G.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-MAR-1996) Roman G. Reddy, Pathobiology, University of  
 Florida, 471 Mowry Rd., Gainesville, FL 32611-0880, USA  
 FEATURES  
 Location/Qualifiers  
 1..1278  
 /organism="Ehrlichia ruminantium"  
 /strain="Um Banein"  
 /db\_xref="taxon:779"  
 396..1268  
 /gene="map1"  
 396..1268  
 /gene="map1"  
 /note="surface protein"  
 /codon\_start=1  
 /transl\_table=11  
 /product="major antigenic protein 1"  
 /protein\_id="AAC44146.1"  
 /db\_xref="GI:1418260"  
 /translation="MNCKKIFITSLISLPLPGVFSVDVQEDSNPAGSVYISAKY  
 MPTASHFGKMSIKEDSKNTQTVGLKMDGDKPSSDSGNSIIFTKGYSFKYENN  
 PFLGAGIAGYMGNGPREFEVSFTFVKNPGKNYKNDAMKCALDTGPGSTOGAT  
 LNSSVWKNENLTIALMLNACVDITLGMPSVPCAGIGTDLVSNATNPKLSYQ  
 CKLGISYSINPASIPIGHHFRVIRVINEFKDITTSKITSKGLATASPGFASATLD  
 VCHGIEIGGRVFF"  
 BASE COUNT 431 a 175 c 207 g 465 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 4,02e-31 Length: 1278  
 Score: 472.50 Matches: 109  
 Percent Similarity: 57.19% Conservative: 62  
 Best Local Similarity: 36.45% Mismatches: 113  
 Query Match: 31.58% Indels: 15  
 DB: 1 Gaps: 8  
 US-10-062-624-40 (1-293) x CRU50835 (1-1278)  
 QY 2 AsnAsnLysLeuLysPheThrIleLeuAenThrValLeuValCysLeuSer---Leu 20  
 Db 393 AATATGAATTCGAAGAAATTTTATCACAAGTACACTAATATCATTCATTTTGA 452  
 QY 21 ProAsnIleSerSerLysLysAlaLeuAsnAsnAlaLysLysTyTyTyGlyLeuTy 40  
 Db 453 CTGTGTGTCCTTTCTCATGTATACAGGAGGATAGCAACCCAGCGGTAGTGTTC 512  
 QY 41 IleSerGlyGlnTyLysProSerValSerValPheSerAsnPheSerValLysGluThr 60  
 Db 513 ATTAGCGCAATATACATGCACTGCTTCACATTTTGGTAAATGTCAATCAAGAGAT 572  
 QY 61 AsnValIleThrLysAsnLeuLeuAlaLeuLysLysAspValAspSerIleGluThrLys 80  
 Db 573 TCAAAAAATACACAAACAGTATTTGGCCTTAAAAAAGATTGGATGGCGTT-----AAA 626

QY 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThr---IleProTyThrAla 99  
 Db 627 ACACCATCATCAGATAGCGGTAGCAATAGTATTTATCTTCTGAAAGACATATTCATC 686  
 QY 100 ValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyThrPheAlaGlu 118  
 Db 687 AATATGAAACACACCCATTTTAGGTTTCTCGAGCAATTTGGGTATTCATG---AAT 743  
 QY 119 GlyThrArgValGluIleGluGlySerTyTyGluGluPheAspValLysAsnProGlyGly 138  
 Db 744 GGTCCAAGATAGAAATTTGAAGTATCTTATGAAACTTTTGTATGTAATAAAACCCAGGTGT 803  
 QY 139 TyThrLeuSerAspAlaTyArgTyPheAlaLeuAlaArgGluMetLysGlyAsnSer 158  
 Db 804 AATTATAAAATGATGACATATGTATTTGTCTCTAGATACAGGAACACACAGGATCT--- 860  
 QY 159 PheThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGly 178  
 Db 861 ---ACTCAGGTGCACACATTAATTCATCTGTT-----ATGGTAAAAAATGAAAAAT 908  
 QY 179 LeuSerIleLeuSerValIleValAsnValCysTyRaspPheSerLeuAsnLeuSer 198  
 Db 909 TTAACGTATATGCACTAATGTTAAATGCATGCTATGATATACACTTTGAAGGAATGCCA 968  
 QY 199 IleSerProTyIleCysGlyValAlaGlyValAspAlaIleGluPheAspValLeu 218  
 Db 969 GTTCTCCATATGTGTGCGAGCATTTGTTACTGATTTAGTGTGATGATCAATCACTGCTACA 1028  
 QY 219 HisIleLysPheAlaTyGlnSerLysLeuGlyIleAlaTySerLeuProSerAsnIle 238  
 Db 1029 AATCCTAAGTTATCTTATCAGGAAAGTTAGTATTAGTATTCAATAAATCTCGAAGCT 1088  
 QY 239 SerLeuPheAlaSerLeuTyTyHisLysValMetGlyAsnGlnPheLysAsnLeuAsn 258  
 Db 1089 TCTATCTTATTGTGGACATTTCCATAGATTTAGTATAGTAATGAATTTAAAGACATTA 1148  
 QY 259 ValGlnHisVal-----AlaGluLeuAlaSerIleProLysIleThrSerAla 274  
 Db 1149 ACTTCCAAAATATTCACCTCACTGTTGTAATAGTACTGCGAGTAGCCAGGTTTGTGCA 1208  
 QY 275 ValAlaThrLeuAsnIleGlyTyPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293  
 Db 1209 TCAGCAACACTTGATGTTTGCCATTCGTTATAGAAATTTGGAGGAAGGTTTGTATT 1265  
 RESULT 15  
 AF368001  
 LOCUS  
 DEFINITION Cowdria ruminantium isolate Burkina Faso major antigenic protein  
 MAP1 gene, complete cds.  
 ACCESSION AF368001  
 VERSION AF368001.1 GI:15429317  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Ehrlichia ruminantium.  
 Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 Anaplasmataceae; Ehrlichia.  
 REFERENCE 1 (bases 1 to 873)  
 AUTHORS Allsopp, M.T., Dorfling, C.M., Maillard, J.-C., Bensaïd, A.,  
 Haydon, D.T., van Heerden, H. and Allsopp, B.A.  
 TITLE Ehrlichia ruminantium major antigenic protein gene (map1) variants  
 are not geographically constrained and show no evidence of having  
 evolved under positive selection pressure  
 J. Clin. Microbiol. 39 (11), 4200-4203 (2001)  
 JOURNAL 21539003  
 MEDLINE 11682561  
 PUBMED  
 REFERENCE 2 (bases 1 to 873)  
 AUTHORS Allsopp, M.T., Dorfling, C., Maillard, J.-C., Bensaïd, A., Van  
 Heerden, H. and Allsopp, B.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-APR-2001) Molecular Biology, Onderstepoort Veterinary  
 Institute, Old Southpansberg Road, Onderstepoort, Gauteng 0110,  
 South Africa  
 FEATURES Location/Qualifiers

QY	214	PhePheAspValLeuHisLeuPheAlaTyrGlnSerLysLeuGlyLeuAlaTyrSer	233
Db	619	GTAAATTAATGCTACAAATCTTAAATATCTTATCAAGGAAAGCTAGGCATAAGTTACTCA	678
QY	234	LeuProSerAsnIleSerLeuPheAlaSerLeuTyrThrHisLysValMetGlyAsnGln	253
Db	679	ATCAATCTCGAAGCTTCTATCTTTATTTGGTGGGCATTTCCATAGATAGATATAGGTAATGAA	738
QY	254	PhyLeuAsnLeuAsnValGlnHisVal-----AlaGluLeuAlaSerIlePro	269
Db	739	TTTAAAGATATTTGCTACCTCAAAATATTTACTTCAACTAGTAAGCAATCATCTATACCT	798
QY	270	LysIleThrSerAlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGluLeuGlyAla	289
Db	799	AATCTGCTTGGCATCAGCAACACTTGTATCTGTCCTCATTTGCGTATAGAAATTTGGAGGA	858
QY	290	ArgLeuThrPhe	293
Db	859	AGGTTGTATTT	870
RESULT	16		
LOCUS	CRU50832		
DEFINITION	Cowdria ruminantium Gardel major antigenic protein 1 (map1) gene, complete cds.		
ACCESSION	U50832		
VERSION	U50832.1	GI:1418253	
KEYWORDS	heartwater rickettsia strain=Gardel.		
SOURCE	Ehrlichia ruminantium		
ORGANISM	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Anaplasmataceae; Ehrlichia.		
REFERENCE	1 (bases 1 to 1278)		
AUTHORS	Reddy, G.R., Sulsona, C.R., Harrison, R.H., Mahan, S.M., Burridge, M.J.		
TITLE	Sequence heterogeneity of the major antigenic protein 1 genes from Cowdria ruminantium isolates from different geographical areas		
JOURNAL	Clin. Diagn. Immunol. 3 (4), 417-422 (1996)		
MEDLINE	96400830		
PUBMED	8807206		
REFERENCE	2 (bases 1 to 1278)		
AUTHORS	Reddy, G.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-MAR-1996) Roman G. Reddy, Pathobiology, University of Florida, 471 Mowry Rd., Gainesville, FL 32611-0880, USA		
FEATURES	Location/Qualifiers		
source	1..1278		
gene	/organism="Ehrlichia ruminantium"		
CDS	/strain="Gardel"		
	/db_xref="taxon:779"		
	396..1268		
	/gene="map1"		
	396..1268		
	/note="surface protein"		
	/codon_start=1		
	/transl_table=11		
	/product="major antigenic protein 1"		
	/protein_id="AAC44143.1"		
	/db_xref="GI:1418254"		
	/translation="MNCKKIFITSLISLVFLPGVFSFSDVIOEDSSPAGSVVISAKY		
	MPTASHFGKMSIKEDSKNTQTVFGLKMDGVKPTSDSNNSIIFTEKDYSPKYN		
	PFLGPAAGIYGVNPGRIEFVSEYETFDVKNPGNYKNDAMHYCALDTCTPGSTQAT		
	LNSVWVKNENLTDIALMNAVDILEGMPVSPYVCAGIGTDLVSVINATNPKLISYQ		
	GKLGISVINPEASIFIGGHHFDIGNEFKDIATSKIFTSTKASSIINPFGFASATLD		
	VCHFGIEIGGRFV"		
BASE COUNT	307 a 139 c 154 g 273 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	3,79e-31	Length:	873
Score:	470.50	Matches:	110
Percent Similarity:	55.92%	Conservative:	60
Best Local Similarity:	36.18%	Mismatches:	109
Query Match:	31.45%	Indels:	25
DB:	1	Gaps:	9
US-10-062-624-40 (1-293) x AF368001 (1-873)			
QY	1	MetAsnLysLeuLysPheThrIleLeuThrValLeuValCysLeuLeuSer---	19
Db	1	ATGAATTCGCAAGAAATTTT-----ATCAAGATACATCAATATCATGTGTCATTT	54
QY	20	LeuProAsnIleSerSerLysLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeu	39
Db	55	TTACCTGGTGTCTCTTCTCATGTATATACAGGAGACAGCCCGGCGAGTGTT	114
QY	40	TyrIleSerGlyClnTyrLysProSerValSerValPheSerAsnPheSerValLysGlu	59
Db	115	TACATTAGCGCAAAATACATGCCAATCGTCATCACATTTTGTAAATGTCAATCAAGAA	174
QY	60	ThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThr	79
Db	175	GATTCAAAATAACTCAACAGATTTTGTCTTAAATAAGATGGGATGGCGTTTAAAGTA	234
QY	80	LysThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIle-----	95
Db	235	CCACATCA-----GAAACAGTAAACAATCTCAAAATTTTACTGAAAAA	279
QY	96	ProTyrThrAlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyr	114
Db	280	GATTATTTCTTCAGATATGAAACAATCCGTTTTTAGGTTTTTGTGGACGAATGGATAC	339
QY	115	ThrPheAlaGluGlyThrArgValGluIleGluGlySerTyrGluGluPheAspValLys	134
Db	340	TCAATG---AATGGGCCAAGATAGAGTTTGAAGTATCTCTATGAAATTTTGTATGTCAA	396
QY	135	AsnProGlyGlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaAaGluMet	154
Db	397	ATCCAGGTGGCACTATAAATAATGATGCACATGCTACTGTGCTTTA-----	444
QY	155	LysGlyAsnSerPheThrProLysGluLysValSerAsnSerIlePheHisThrValMet	174
Db	445	-----GAGACAGCAGCAGCACTAGTAATCAAGGGGCAACATTAGCTTCATCTGTTATG	498
QY	175	---ArgAsnAspGlyLeuSerIleLeuSerValIleValAsnValCysTyrAspPheSer	193
Db	499	GTAATAAACGAAATTTTACAGATATCATTAATGTTAAATGTCATGTTATGATATAATG	558
QY	194	LeuAsnAsnLeuSerIleSerProTyrIleCysGlyAlaGlyValAspAlaIleGlu	213
Db	559	CTTGATGGAAATGCCAGTTTCTCCATGTATGTGCGAGTATTTGGTACTGATTAGTATCA	618

Best Local Similarity: 36.12% Mismatches: 113  
Query Match: 31.38% Indels: 15  
DB: 1 Gaps: 8

US-10-062-624-40 (1-293) x CRU50832 (1-1278)

Qy 2 AnAsnLysLeuLysPheThrIleLeuValLeuValCysLeuLeuSer---Leu 20  
Db 393 AATATGAATTCGACGAAATTTTATCAGAGTACACTAATATCATTTAGTGTGCTTTTA 452

Qy 21 ProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysLeuTyrrGlyLeuTyrr 40  
Db 453 CTTGGTGTCTTTTCTGATGTAATACAGGAAGATAGCAGCCAGCAGGTAGTCTTTAC 512

Qy 41 IleSerGlyGlnTyrrLysPheSerValSerValPheSerAsnPheSerValLysGluThr 60  
Db 513 ATTAGCGCAAAATACATGCATCACTGCATCACTATTTTGGTAAATGCTCAATCAAGAAGAT 572

Qy 61 AnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80  
Db 573 TCAAAAATACCTCAACAGATTTTGGCTTAAAAAAGATGGGATGGCTT-----AAA 626

Qy 81 ThrAspAlaSerValGlyLysSerAsnProSerAsnPheThr---IleProTyrrThrAla 99  
Db 627 ACACCATCATCATAGAGCGGTAAACATAGTATTATCTTCACCTGAAAAAGACTATTCAATTC 686

Qy 100 ValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrrThrPheAlaGlu 118  
Db 687 AATATGAAACAAACCCATTTTGTAGTTTGTGGAGCAATTTGGGTATTCAATG---AAT 743

Qy 119 GlyThrArgValGluIleGlySerTyrrGluGluPheAspValLysAsnProGlyGly 138  
Db 744 GGTCCAGAAATAGATTTGAGTACTTATGAACTTTTGAATGATTAATAAACCCAGGTGT 803

Qy 139 TyrrThrLeuSerAspAlaTyrrArgTyrrPheAlaLeuAlaArgGluMetLysGlyAsnSer 158  
Db 804 AATTATAAAATGATGCATATGATTTGTCTGTAGTACAGGAACACCGAGTCT--- 860

Qy 159 PheThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGly 178  
Db 861 ---ACTCAGGTGCAACATTAATTCATCTGTT-----ATGTTAAATAATGAAAT 908

Qy 179 LeuSerIleLeuSerValIleValAsnValCysTyrrAspPheSerLeuAsnAsnLeuSer 198  
Db 909 TTAATGATATTTGCACTATGTAATGATGATGATGATGATGATGATGATGATGATGATGAT 968

Qy 199 IleSerProTyrrIleCysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeu 218  
Db 969 GTTCTCCATATGTTGTGCGGCATTTGTTAGTATTTAGTATTTAGTATTTAGTATTTAGT 1028

Qy 219 HisIleLysPheAlaTyrrGlnSerLysLeuGlyIleAlaTyrrSerLeuProSerAsnIle 238  
Db 1029 AATCCTAAGTTATCTTATCAGGAAAGTTAGTATTTAGTATTTAGTATTTAGTATTTAGT 1088

Qy 239 SerLeuPheAlaSerLeuTyrrTyrrHisLysValMetGlyAsnGlnPheLysAsnLeuAsn 258  
Db 1089 TCTATCTTTATTTGGTGGACATTTCCATAGATTTAGTATTTAGTATTTAGTATTTAGTATTT 1148

Qy 259 ValGlnHisVal-----AlaGluLeuAlaSerIleProLysIleThrSerAla 274  
Db 1149 ACTTCCAAAATATTACCTCAACTGTTAAATTTAGTCTAGTCTAGTCTAGTCTAGTCTAGT 1208

Qy 275 ValAlaThrLeuAsnIleGlyTyrrPheGlyGlyGluIleGlyAlaArgLeuThrPhe 293  
Db 1209 TCAGCAACACCTTGATGTTTGGCATTTCCGTATAGAAATTTGGAGAGGTTTGTATTT 1265

RESULT 17  
AX042314  
LOCUS AX042314 843 bp DNA linear PAT 23-NOV-2000  
DEFINITION Sequence 10 from Patent WO0065063.  
ACCESSION AX042314  
VERSION AX042314.1 GI:11340974  
KEYWORDS

SOURCE Ehrlichia chaffeensis.  
ORGANISM Ehrlichia chaffeensis  
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
Anaplasmataceae; Ehrlichia.  
REFERENCE 1 (bases 1 to 843)  
AUTHORS Barbet, A.F., Bowie, M.V., Cantu, R.R., Burridge, M.J., Mahan, S.M.,  
McQuire, T.C., Rurangirwa, P.R., Moreland, A.L., Simbi, B.H.,  
Whitmire, W.W. and Alleman, A.R.  
TITLE Nucleic acid vaccines against rickettsial diseases and methods of  
use  
JOURNAL Patent: WO 0065063-A 10 02-NOV-2000;  
UNIVERSITY OF FLORIDA (US)  
FEATURES Location/Qualifiers  
source 1..843  
/organism="Ehrlichia chaffeensis"  
/db\_xref="taxon:945"  
BASE COUNT 282 a 127 c 159 g 275 t  
ORIGIN  
Alignment Scores: 5.37e-31 Length: 843  
Pred. No.: 468.50 Matches: 115  
Score: 56.03% Conservative: 57  
Percent Similarity: 37.46% Mismatches: 94  
Best Local Similarity: 31.32% Indels: 41  
Query Match: 6 Gaps: 14  
DB:  
US-10-062-624-40 (1-293) x AX042314 (1-843)  
Qy 1 MetAsnLysLeuLysPheThrIleAsnThrValLeuValCysLeuLeuSer--- 19  
Db 3 ATGAATTCGAAA---AAATTT---TTTATAACAACATACATTAGTATCGTAAATGCTCTTC 56  
Qy 20 LeuProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyrrTyrrGlyLeu 39  
Db 57 TTACCTCGAATATCATTTTCTGATGCGATGACAGAACGAC---AATGTTGGTGGTAATTC 113  
Qy 40 TyrrIleSerGlyGlnTyrrLysProSerValSerValPheSerAsnPheSerValLysGlu 59  
Db 114 TATATCATGTTGGAATATGATGACCAAGTGTTCACATTTTGGCGTATTTCTGCTAAACAG 173  
Qy 60 ThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThr 79  
Db 174 GAAAGAAATACAAACATCGGAGTATTTGGATTAAAGCAAGATTTGGATGGC----- 224  
Qy 80 LysThrAspAlaSerValGlyLysSerAsnProSerAsn---PheThrIlePro---Tyr 97  
Db 225 -----AGCAACAATATCTAAATAATCTCCAGAAATACATTTAAACGTTCCAAATAT 275  
Qy 98 ThrAlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrrThrPhe 116  
Db 276 TCATTTAAATATGAATAATCCATTTCTAGTTTTCAGGAGCTGTGGTTAT---TTA 332  
Qy 117 AlaGluGlyThrArgValGluIleGluGlySerTyrrGluGluPheAspValLysAsnPro 136  
Db 333 ATGAATGTTCCAGAAATAGAGTTAGAAATGCTTATGAAACATTTGATGTGAAACACAG 392  
Qy 137 GlyGlyTyrrThrLeuSerAspAlaTyrrArgTyrrPheAlaLeuAlaArgGluMetLysGly 156  
Db 393 GGTAATAACTATAAGACGATGCTCACAATAATATTATGCTTTTAAACCCATAACAGTGGGGA 452  
Qy 157 AsnSerPheThrProLysGluLysValSerAsnSer-----IlePheHisThr 172  
Db 453 -----AAGTAGCAATCAGGTGATAGTTGTTT----- 485  
Qy 173 ValMetArgAsnAspGlyLeuSerIleLeuSerValIleValAsnValCysTyrrAspPhe 192  
Db 486 ---CTAAAAAATGAAGACTACTTGTATATACACTTATGTTGTTGTTGTTGTTGTTGTTGTT 542  
Qy 193 SerLeuAsnLeuSerIleSerProTyrrIleCysGlyGlyAlaGlyValAspAlaIle 212  
Db 543 ATAAGTGAAGGAATACCTTTCTCTCTTACATATGTTGCGAGGTGTTGGTACTGATTTAATA 602



QY 213 GluPhePheAspValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyr 232  
 Db 603 TCCATGTTTGAAGCTATAAACCTAAAAATTTCTTATCAAGGAATAGGTTTCAAGTTAC 662  
 QY 233 SerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsn 252  
 Db 663 TCCATAAGCCACGAGCTTCTGTTTGTGGACATTTTCATAGGTGATAGGGAAT 722  
 QY 253 GlnPheLysAsnLeuAsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThr 272  
 Db 723 GAATTCAGAGATATTCCT-----AlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGlu 286  
 QY 273 Ser-----AlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGlu 286  
 Db 762 ACTCTCAGAGTAATCACTTTACTATAGTAACACTAAGTATGATGCCACTTTGGAGTGGA 821  
 QY 287 IleGlyAlaArgLeuThrPhe 293  
 Db 822 CTGGAGGAGGTTTAACTTT 842

RESULT 18  
 AF062761  
 LOCUS  
 DEFINITION Ehrlichia chaffeensis 28 kDa major surface antigen multi-gene  
 locus, partial sequence.  
 ACCESSION AF062761  
 VERSION AF062761.1 GI:3327958  
 KEYWORDS  
 SOURCE  
 ORGANISM Ehrlichia chaffeensis.  
 Ehrlichia chaffeensis  
 Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 Anaplasmataceae; Ehrlichia.  
 1 (bases 1 to 4683)  
 Reddy,G.R., Sulsona,C.R., Barbet,A.F., Mahan,S.M., Burridge,M.J.  
 and Alleman,A.R.  
 TITLE Molecular characterization of a 28 kDa surface antigen gene family  
 of the tribe Ehrlichiae  
 Biochem. Biophys. Res. Commun. 247 (3), 636-643 (1998)  
 JOURNAL  
 MEDLINE 98321180  
 PUBMED 9647746  
 REFERENCE 2 (bases 1 to 4683)  
 Reddy,G.R., Sulsona,C.R., Barbet,A.F., Mahan,S.M., Burridge,M.J.  
 and Alleman,A.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-MAY-1998) Diagnostic Medicine/Pathobiology, College  
 of Veterinary Medicine, Kansas State University, 1800 Denison Ave,  
 Manhattan, KS 66506, USA

FEATURES  
 source  
 1..4683  
 /organism="Ehrlichia chaffeensis"  
 /strain="Arkansas"  
 /db\_xref="taxon:945"  
 <1..134  
 /note="similar to Cowdria ruminantium MAP1 gene; ORF 1;  
 silent copy"  
 /codon\_start=3  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="28 kDa major surface antigen-1"  
 /protein\_id="AAC26717.1"  
 /db\_xref="GI:3327960"  
 /translation="NEFRDITSLKAPATPPSAATPDLATVTLVSVCHFVGLGGRNF"  
 179..188  
 /note="palindromic sequence part 1"  
 /evidence=not\_experimental  
 203..212  
 /note="palindromic sequence part 2"  
 /evidence=not\_experimental  
 226..239  
 /note="g-rich sequence"  
 349..354  
 /evidence=not\_experimental  
 375..380

CDS  
 1..434  
 /evidence=not\_experimental  
 430..434  
 /evidence=not\_experimental  
 439..1299  
 /note="similar to Cowdria ruminantium MAP1 gene; ORF 2;  
 silent copy"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="28 kDa major surface antigen-2"  
 /protein\_id="AAC26718.1"  
 /db\_xref="GI:3327961"  
 /translation="MNCKEFFITTLVLSMSFLPGISLSDPVDQDNISGNFYISGKYM  
 PSASHFGVFSKAEERTTVGVFGIEQMDRCVSRITLSDIFTVPNTSFYKNNLPSG  
 FAGAIGYSDMGPRILEVSYEAFDVKQGNKYKNEAHRYVALSHLLGTETQIDAGSA  
 SVFLINGLDKSKFLNACVYDSEGIPFSPYICAGVGTDLISMFATNPISYQKGLS  
 KSGELDILPMLNACVYDSEGIPFSPYICAGVGTDLISMFATNPISYQKGLS  
 SINPEASVFIGGHFKVIGNEFRDITPLKAFVTSSATPDLAIVTLVSVCHFVGLGGRF  
 NF"  
 2444..2456  
 /note="palindromic sequence part 1"  
 /evidence=not\_experimental  
 2465..2477  
 /note="palindromic sequence part 2"  
 /evidence=not\_experimental  
 2491..2505  
 /note="g-rich sequence"  
 2618..2624  
 /evidence=not\_experimental  
 2644..2649  
 /evidence=not\_experimental  
 2699..2703  
 /evidence=not\_experimental  
 2708..3550  
 /note="similar to Cowdria ruminantium MAP1 gene; ORF 4;  
 silent copy"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="28 kDa major surface antigen-4"  
 /protein\_id="AAC26720.1"  
 /db\_xref="GI:3327963"  
 /translation="MNCKEFFITTLVLSMSFLPGISLSDPVDQDNISGNFYISGKYM  
 PSASHFGVFSKAEERTTVGVGLKQDWDGSTISKNSPENTFNPNTSFYKNNLPSG  
 FAGAVGLMNGPRIELMSYETFDVKQGNKYKNEAHRYVALTHNSGGKLSNAGDKFY

terminator  
 terminator  
 misc\_feature  
 -35\_signal  
 -10\_signal  
 RBS  
 CDS



KGKLSISINSEASIFGGHPRHVRIGNEFKDIATLKLFTATNKVSTVANPGFASATLD  
VCHFGIRIGGRFIF"

BASE COUNT 525 a 216 c 251 g 572 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1,35e-30 Length: 1564  
Score: 467.50 Matches: 108  
Percent Similarity: 57.81% Conservativeness: 66  
Best Local Similarity: 35.88% Mismatches: 108  
Query Match: 31.25% Indels: 19  
DB: 1 Gaps: 9

US-10-062-624-40 (1-293) x CRU49843 (1-1564)

Qy 2 AsnAsnLysLeuLysPheThrIleAsnThrValLeuValCysLeuLeuSer---Leu 20  
Db 607 AATATGATTCGAAGAAATTTTATCACAAGTACACTAATATCATGTAGTGTATTTA 666  
Qy 21 ProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40  
Db 667 CCGTGTGTCTCTTCTGATGTAATACAGGAGATAGCAACCCAGCTGGTAGTGTATAC 726  
Qy 41 IleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60-  
Db 727 ATTAGCGCAAAATACATGCCAACCTGCTTCACACTTTGGTAAATGTCAATTAAGAAGAT 786  
Qy 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80  
Db 787 TCAGAAAATACCTCAACAGCTGTTGGICTAAAAAAGATTGGGATGGATTAAAGTACCA 846  
Qy 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThr---IleProTyrThrAla 99  
Db 847 ACATCAGAAAAACCAAACTACTCT-----TCACCTTTTACTGAAAAAGATTATCTTC 900  
Qy 100 ValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGlu 118  
Db 901 AGATATGAAACAAATCCGTTTTAGTTTCTGGAGCAATTTGGGTACTCAATG---AAT 957  
Qy 119 GlyThrArgValGluIleGluGlySerTyrGluGluPheAspValLysAsnProGlyGly 138  
Db 958 GGACCAAGAATAGAGTTCGAAGTATCTCTAGAACTTTGTATGTAATAAAACCCAGGTGC 1017  
Qy 139 TyrThrLeuSerAspAlaTyrArgTyrPheAlaLeu-----AlaArgGluMetLysGly 156  
Db 1018 AACTACAAAAATGATGCACATGCTACTGCTTTAGATACACACACAGAGTGTACT 1077  
Qy 157 AsnSerPheThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsn 176  
Db 1078 AATGGTGCACAC-----TTAGCTTCATCTGTT-----ATGATAAAAAAT 1116  
Qy 177 AspGlyLeuSerIleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsn 196  
Db 1117 GAAATTTACAAATATATCATTAATGTAATCGGTGTATGATATATATGCTTGATGGA 1176  
Qy 197 LeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPheAsp 216  
Db 1177 ATGCCAGTTCTCCATATGATGTGCAGGTATTGCTAGTATTAGTGTAGTAAATTAAT 1236  
Qy 217 ValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSer 236  
Db 1237 GCTCAAAATCTTAATATCTTATCAAGGAAGTAGGCATTAAGTAACTCAATCAATCT 1296  
Qy 237 AsnIleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGlnPheLysAsn 256  
Db 1297 GAAGCTTCTACTTATTCGGTGGACATTCATAGATTATAGTAAATTAAGAT 1356  
Qy 257 LeuAsnValGlnHisVal-----AlaGluLeuAlaSerIleProLysIleThr 272  
Db 1357 ATTGCTACTTTAAAAATATTTACCGCAACTAATAAGTATCTACTGTAGCCCAATCCAGGT 1416  
Qy 273 SerAlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGluIleGlyAlaArgLeuThr 292

Db 1417 TTTCATCAGCAACACTTGATGTGTGTCTACTTTTGTATAGAAATGGAGGAGTTTATA 1476  
Qy 293 Phe 293  
Db 1477 TTT 1479  
RESULT 20  
AF068234/c 2977 bp DNA linear BCT 24-MAY-2000  
LOCUS Ehrlichia chaffeensis 28 kDa outer membrane protein gene cluster,  
DEFINITION partial sequence.  
ACCESSION AF068234 AF230643  
VERSION AF068234.2 GI:8052524  
KEYWORDS Ehrlichia chaffeensis.  
SOURCE Ehrlichia chaffeensis.  
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
Anaplasmataceae; Ehrlichia.  
REFERENCE 1 (bases 1 to 2977)  
AUTHORS Yu,X.J., McBride,J.W. and Walker,D.H.  
TITLE Genetic diversity of the 28-kilodalton outer membrane protein gene  
in human isolates of Ehrlichia chaffeensis  
J. Clin. Microbiol. 37 (4), 1137-1143 (1999)  
JOURNAL 99175287  
MEDLINE 10074538  
REFERENCE 2 (bases 1 to 2977)  
AUTHORS Yu,X.J., McBride,J.W., Zhang,X. and Walker,D.H.  
TITLE Characterization of the complete transcriptionally active Ehrlichia  
chaffeensis 28 kDa outer membrane protein multigene family  
Gene 248 (1-2), 29-68 (2000)  
JOURNAL 99175287  
REFERENCE 3 (bases 1 to 1268)  
AUTHORS Yu,X.J., McBride,J.W. and Walker,D.H.  
TITLE Direct Submission  
JOURNAL Submitted (26-MAY-1998) Pathology, The University of Texas Medical  
Branch, 301 Univ. Blvd., Galveston, TX 77555, USA  
REFERENCE 4 (bases 1 to 2977)  
AUTHORS Yu,X.J., McBride,J.W., Zhang,X. and Walker,D.H.  
TITLE Direct Submission  
JOURNAL Submitted (01-FEB-2000) Pathology, University of Texas Medical  
Branch, 301 Univ. Blvd., Galveston, TX 77555, USA  
REMARK Sequence update by submitter  
COMMENT On May 24, 2000 this sequence version replaced gi:3192922.  
FEATURES  
source 1. 2977  
/organism="Ehrlichia chaffeensis"  
/strain="Arkansas"  
/db\_xref="taxon:945"  
complement (180. .995)  
/genes="p28-20"  
complement (180. .995)  
/genes="p28-20"  
/notes="component of 28-kDa outer membrane protein  
multigene family"  
/codon\_start=1  
/transl\_table=11  
/product="p28-20 precursor"  
/protein\_id="AAC19134.2"  
/db\_xref="GI:8052525"  
/translations="MNYKFKVGVVALTLLSLPDNSFSDANVPGEKGFVCTQYKV  
GVNFNFSABETLPGLTKSIFALGDKSSISDHAGFTQAYNTYASNPAGFGVIGY  
YVDFRVEFEGAYENFEPQWYPEGESHKFFALRSBSTVQDNKFLVLENDGVIDKS  
LNVFCYDIAHSGISIPLAYMCAGVADYIKFLGISLPKFSYQKFGVNYFVSNVMLF  
GGGYHKVIGIYRVERVEIAYHPATLTNPVKTTSASATLDTDFGWEVGMFTL"  
sig\_peptide complement (921. .995)  
mat\_peptide complement (183. .920)  
/genes="p28-20"  
complement (1178. .2014)  
/genes="p28-21"  
complement (1178. .2014)  
/genes="p28-21"  
/notes="component of 28-kDa outer membrane protein

```

multigene family"
/codon_start=1
/transl_table=1
/product="p28-21"
/protein_id="AAF71791.1"
/db_xref="GI:8052526"
/translation="WRYRISVILMLLVPCCFSGSLDVSLSNRLKPVFLGTSYK
LSAPLFSFGISGETTRINGKQVDRVGLKSDILLDDAKMDKDFNNFNSESEVXPYKYN
NIFGUSFIFGYSRLRVELEGSYKFDVIDTRNLHVDNNYRHIALVRSNPTLYDYF
VLKNDGVFYFILNICYDVAVDNIVFSCVIGEDIIKIPDSIRFPFSNKLGIN
YLMQDMLLFDDVYHRYVGVNENYNNIPQYVSLPNPLNISTAAKLDMFYFGAIEIGI
FV"
BASE COUNT 1172 a 529 c 353 g 923 t
ORIGIN

Alignment Scores:
Pred. No.: 5,7e-30 Length: 2977
Score: 464.00 Matches: 115
Percent Similarity: 54.67% Conservative: 49
Best Local Similarity: 38.33% Mismatches: 98
Query Match: 31.02% Indels: 38
DB: 1 Gaps: 12

US-10-062-624-40 (1-293) x AF068234 (1-2977)

Qy 1 MetAsnLeuValLeuValPheThrIleAsnThrValLeuValCysLeuLeuSer--- 19
Db 995 ATGAATTACAAA---AAATTT---GTTGAGGTGTTGCATTGGCTACATTGCTTTCTTC 942
Qy 20 LeuProAsnLeuSerSerLysAlaIleAsnAsnAlaLysLysTyrrGlyLeu 39
Db 941 TTGCTGATAATCTTTTCTGATGCAACGTTCTCTGAAGGGAAG-----GGATT 888
Qy 40 TyrIleSerGlyGlnTyrLysProSerValPheSerAsnPheSerValLysGlu 59
Db 887 TATGAGGTACTCAATATAAAGTGTGTCTTAATTTTCAGTAATTTTTCAGCTGAAGA 828
Qy 60 Thr---AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGlu 78
Db 827 ACACCTCTCGGACTCACAAAAGTATTTTGTGTTAGGCTTTGATAAATCTAGC----- 774
Qy 79 ThrLysThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyrThr 98
Db 773 -----ATTAGTGCATGTCAGGTTTATACACACCGTATATAT 738
Qy 99 AlaValPheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGlu 118
Db 737 CCTACATATGCAAGCAATTTTCTGGGTTTGTGGTGTATTGGATAT---TATGTTAAT 681
Qy 119 GlyThrArgValGluIleGluGlySerTyrGluGluPheAspValLysAsn----- 135
Db 680 GATTTTAGAGTAGAGTTTGAGGTGCTTATGAAAATTTTGAACCCGAAAGACAGTGGTAT 621
Qy 136 ProGlyGlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGlu----- 153
Db 620 CTGGAAGG-----GGAGAAGCCATAAGTTTTTTCCTATCTCGTAGGCTCACT 570
Qy 154 MetLysGlyAsnSerPheThrProLysGluLysValSerAsnSerIlePheHisThrVal 173
Db 569 GTTCAAGATAATAAGTTT-----ATAGTG 546
Qy 174 MetArgAsnAspGlyLeuSerIleLeuSerValIleValAsnValCysTyrAspPheSer 193
Db 545 TTAGAAAATGATGGTGTATTATGATAAATCTCTTAATGTGAACCTTTGTTATGATATTC 486
Qy 194 LeuAsnAsnLeuSerIleSerProTyrIleCysGlyValAlaGlyValAspAlaIleGlu 213
Db 485 CATGTTAGTATTCCTTTAGCACCATATATGTGTGCTGGTGTGGAGCAGATTACATAAAA 426
Qy 214 PhePheAspValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSer 233
Db 425 TTTTGGGAATATCATTTGCTAAATTTCTTATCAAGTGAAATTTGGGGTTAACTATCTCT 366

```

234 LeuProSerAsnLeuSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsnGln 253  
 365 GTGAGTGTAAATGTTATTTGTTGGAGGTTATTATCATATAAGGTTATAGGTAATAGG 306  
 254 PheLysAsnLeuAsnVal---GlnHisValAlaGluLeuAlaSerIleProLysIleThr 272  
 305 TATGAAGAGTAGAAATAGCTTATCATCTGCCACATTACCAATGTGCTTAAACTACT 246  
 273 SerAlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGluIleGlyAlaArgLeuThr 292  
 245 TCAGCATCTGCTACTTATAGATACTGATTATTTTGGTGGGAAGTTGGTATGAGATTACA 186

RESULT 21  
 ECU72291/c 27190 bp DNA linear BCT 02-APR-2001  
 LOCUS Ehrlichia chaffeensis strain Arkansas major outer membrane protein  
 DEFINITION Omp-1 multigene cluster, complete sequence.  
 ACCESSION U72291 AF021338  
 VERSION U72291.2 GI:13511827  
 KEYWORDS Ehrlichia chaffeensis.  
 SOURCE Ehrlichia chaffeensis.  
 ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 Anaplasmataceae; Ehrlichia.  
 1 (bases 14844 to 21136; 21479 to 22234)  
 Ohashi,N., Zhi,N., Zhang,Y. and Rikihisa,Y.  
 Immunodominant major outer membrane proteins of Ehrlichia  
 chaffeensis are encoded by a polymorphic multigene family  
 Infect. Immun. 66 (1), 132-139 (1998)  
 98084465  
 9423849  
 2 (bases 1 to 27190)  
 Ohashi,N., Rikihisa,Y. and Unver,A.  
 Analysis of transcriptionally active gene clusters of major outer  
 membrane protein multigene family in Ehrlichia canis and E.  
 chaffeensis  
 Infect. Immun. 69 (4), 2083-2091 (2001)  
 21153566  
 11254561  
 3 (bases 1 to 27190)  
 Ohashi,N., Zhi,N., Zhang,Y. and Rikihisa,Y.  
 Direct Submission  
 Submitted (24-SEP-1996) Department of Veterinary Biosciences, The  
 Ohio State University, 1925 Coffey Rd., Columbus, OH 43210-1093,  
 USA  
 4 (bases 1 to 27190)  
 Ohashi,N., Zhi,N., Zhang,Y. and Rikihisa,Y.  
 Direct Submission  
 Submitted (27-AUG-1997) Department of Veterinary Biosciences, The  
 Ohio State University, 1925 Coffey Rd., Columbus, OH 43210-1093,  
 USA  
 5 (bases 1 to 27190)  
 Ohashi,N., Rikihisa,Y. and Unver,A.  
 Direct Submission  
 Submitted (29-NOV-2000) Department of Veterinary Biosciences, The  
 Ohio State University, 1925 Coffey Rd., Columbus, OH 43210-1093,  
 USA  
 REMARK Sequence update by submitter  
 COMMENT On or before Apr 2, 2001 this sequence version replaced gi:2853584,  
 gi:2853273.  
 FEATURES  
 source  
 1..27190 Location/Qualifiers  
 /organism="Ehrlichia chaffeensis".  
 /strain="Arkansas"  
 /db\_xref="taxon:945"  
 315..956  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical transcriptional regulator"  
 /protein\_id="AAC02808.2"  
 /db\_xref="GI:13511828"  
 /translation="MSTHAKNKINKMEDTDEIKAHDPVDECVGKEIKRQIRMRGMSQNLANKLG  
 DDHSTKNFTKEGVESTKDLKNSITTKIRPHVPDECVGKEIKRQIRMRGMSQNLANKLG

gene  
CDS  
ITFQVQYKGTNRIVISRLYQLASVLNVEVRDIMLKQEDLKNISCDHTVASTSTL  
KDNEEKIPEPHFDKIDSKSKEVLMVMRAYTCTIKNEKVRNIIYNLVKALSLDNQK"  
1136..2032  
/gene="omp-1m"  
1136..2032  
/gene="omp-1m"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1m"  
/protein\_id="AAK28658.1"  
/db\_xref="GI:13511829"  
/translation="MILINMKLFYHLNIMTKPSAIGIVLSLVTLPACNVFASPIDP  
FNSBETAGFVASHGYNIQPFPSPIVVKYKIDNTENKELTFLTKRESTDPTPKQKS  
EFNDKGYSPLYNRYTGFCAIGYSGGLRVELEGAFTRFDVKQKDKNDYRFAL  
CKDQSIESTNSNGHVMKNEGFRVLSLTFNACYDMIVSNSLSLVPACIGIGGINT  
PLGNTNHTLPLKALGLGLFSLPKTVIFANGYYVKTQKNSFTLSVQVPLEKCAPKH  
IDPIACFNADNYGGEVGLRFL"  
2498..3352  
/gene="omp-1n"  
2498..3352  
/gene="omp-1n"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1n"  
/protein\_id="AAK28659.1"  
/db\_xref="GI:13511830"

gene  
CDS  
/translation="MSKRSNRKFLVWMLILFTPHISLASVLDHNSMVYIOVKPAR  
OHLKSLIKESAANTVVRGLKDLLNDLLTGIDNTNFAIKYNPYENNRLGSGIF  
GYNNKFRIESELSYTFHKNNGYKRICERKALAKELSGSSNPANKYVITLIN  
NGISLTALINVCYDVGDLKHNIITYISGLGVDITIDFLSKYTTKFSYQKLGASYV  
SPQSVPIEGYHGLFGKPEKIPVNPYCDYPSPTPNSKRPVHTTALAMLSIGYGG  
SIGIKPIL"  
3459..3905  
/codon\_start=1  
/transl\_table=11  
/product="unknown function UN2"  
/protein\_id="AAK28660.1"  
/db\_xref="GI:13511831"  
/translation="MNNKSRGOVLLFAAYSIFKYINSCTSKPLRSVKCFYNPLKPY  
YKSTFLKGLKFFVNNYSKTVLSHSLKELGLIITNFKLSYLNFTNPRVLIIY  
YKILNKNSSINLLNKNTIINTATSFINSVLIIGYGFVNINY"  
3915..4808  
/gene="omp-1q"  
3915..4808  
/gene="omp-1q"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1q"  
/protein\_id="AAK28661.1"  
/db\_xref="GI:13511832"

gene  
CDS  
/translation="MSYAKVFILICLLVPSLSFAVNNDFLKNIGHYIGQYKP  
GVPRNPLVNNIRLMSDECRITPHMVQSVQAGTLPPEALEELAKGLLHGIGY  
LFFTLPNYPTKLLAGGIVSYSTHFRVEBAFYKFNLPAPAGYLNKFNVEYFA  
LATMTDKPHQSAEDYYTKMNTGITLSPFIINACYDFILKTRNVAPYLCIGVGN  
FIDPLDQSPKFAQKYGISYFVSPNIAFFIDGSPHGLNNOFSDLPVVDYSSGPP  
TISAKNANFLTSSIGIRFIS"  
4825..5682  
/gene="omp-1p"  
4825..5682  
/gene="omp-1p"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1p"  
/protein\_id="AAK28662.1"  
/db\_xref="GI:13511833"  
/translation="MQKLYISFIIISGLLLPKYVFCMHQNNIDGSYVTKYOLTTPH  
KKNFYIKETDFDCEPGLAKITANTKFDTLKENFSPLHQDSDYSQNDLLIGL  
SVGLFVKSFRIEFGAYKNFTKLARYKSKDGKYKPAIPRKSHGFLDNTFGYTVAK  
NNGISIIISINILCSCTKYKAFPTYICIGVGDPFIEFDVMRIKFAYGQKGVSYPT

gene  
CDS  
/translation="MTKKNFNVNLTFLFLPLKSFYANNNTITOKVGLVISQ  
YKSIPIHPKNFSVEENDKVDLGLTDDVTYTEHLRDNKTNHTVIAKPNFINP  
SSAIGYSGGPRLEIESSYGDVDVNNYKNAVQDNNRYFALVREKNGSNFSPKPHET  
SQSDSNPKSGFYTLMKNGVFWASVINGCYDFSNNTTISPVVCIGVGDDTFEPE  
VMHIFACOSKGYISYPIPSITIFADAHYKVINNFNNLHVYKYSYELKNSPTTISA  
TAKLNIETFGGVEGMRFP"  
7438..8277  
/gene="omp-1v"  
7438..8277  
/gene="omp-1v"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1v"  
/protein\_id="AAK28665.1"  
/db\_xref="GI:13511836"  
/translation="MSKKFFITITGLASLLSFLSIESFSAINHNHTGNTSGIYITG  
QYRPGVHFSNFSVKETNVDITQLGVYKKSASSIDPNTVSNFQGPYTVTFQDNASFS  
GAIGYSVPSRLRLESGSEKFDVKDQYSAKDAFRFALARTSTTVTPDAQKYTM  
KNNGLSVASIMINGCYDLSENNAVSPYICAGIGEDFIEFDLHLIKLAYQKGLGISY  
YFPFKINVPAGYHHRVIGNKFNLNHNVTLUDEFPKATSAVATLNVATFGEAGVK  
FTP"  
8300..8890  
/codon\_start=1  
/transl\_table=11  
/product="unknown function UN3"  
/protein\_id="AAK28666.1"  
/db\_xref="GI:13511837"  
/translation="MIYKEKLRVGEYILAYLSFILSTYIFLVLNIRNYSLAICVI  
SLRITFNVTSTKLLKDKCRDTKFSNMCYLGKPLNLQIFYGIFSFIRNFOUNTLI  
IPNDSKCGFTVTLWDPALHYTVTLTGSEVRNFPDILYENIICQCKLLINYSRVLNQ  
HNKNTLVIIPIPNAREFSNIRVRNISINKESSEYEC"  
8880..9731  
/gene="omp-1w"  
8880..9731  
/gene="omp-1w"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1w"  
/protein\_id="AAK28667.1"

gene  
CDS  
/translation="MIYKEKLRVGEYILAYLSFILSTYIFLVLNIRNYSLAICVI  
SLRITFNVTSTKLLKDKCRDTKFSNMCYLGKPLNLQIFYGIFSFIRNFOUNTLI  
IPNDSKCGFTVTLWDPALHYTVTLTGSEVRNFPDILYENIICQCKLLINYSRVLNQ  
HNKNTLVIIPIPNAREFSNIRVRNISINKESSEYEC"  
8880..9731  
/gene="omp-1w"  
8880..9731  
/gene="omp-1w"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1w"  
/protein\_id="AAK28667.1"

Alignment Scores:  
Pred. No.: 7,71e-29  
Score: 464.00  
Percent Similarity: 54.67%

gene  
CDS  
SKULSINGYHKVIGNKFFELLPYQPVVELKRLVTNKTSKDIDQDVTSATLNLNLEHFS  
SEIGLGFIP"  
5695..6513  
/gene="omp-1t"  
5695..6513  
/gene="omp-1t"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1t"  
/protein\_id="AAK28663.1"  
/db\_xref="GI:13511834"  
/translation="MTMYNKKHYCIYVYVITLFFLLPIESLSALIGNVEKOLKVS  
TYVSYQKPSIFHRNFSIQEHPKSKSESEFKIKANLNILKSNAYNLQFDNNTFP  
SGTIGYFSGKRLAEAGCYQFNKSNNSLIISNNKHSRIHENDYAITTNKLSIA  
SIWNTCYDISINNTSIVPYLCIGEDLVGLFNTIHFKLAYQCKVGMSYLLNNILL  
FSDIYHKVMGNRFKNLYMQYVADPNISEETIPILAKDIGFSEIGIRMFEN"  
6529..7416  
/gene="omp-1u"  
6529..7416  
/gene="omp-1u"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1u"  
/protein\_id="AAK28664.1"  
/db\_xref="GI:13511835"  
/translation="MTKKNFNVNLTFLFLPLKSFYANNNTITOKVGLVISQ  
YKSIPIHPKNFSVEENDKVDLGLTDDVTYTEHLRDNKTNHTVIAKPNFINP  
SSAIGYSGGPRLEIESSYGDVDVNNYKNAVQDNNRYFALVREKNGSNFSPKPHET  
SQSDSNPKSGFYTLMKNGVFWASVINGCYDFSNNTTISPVVCIGVGDDTFEPE  
VMHIFACOSKGYISYPIPSITIFADAHYKVINNFNNLHVYKYSYELKNSPTTISA  
TAKLNIETFGGVEGMRFP"  
7438..8277  
/gene="omp-1v"  
7438..8277  
/gene="omp-1v"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1v"  
/protein\_id="AAK28665.1"  
/db\_xref="GI:13511836"  
/translation="MSKKFFITITGLASLLSFLSIESFSAINHNHTGNTSGIYITG  
QYRPGVHFSNFSVKETNVDITQLGVYKKSASSIDPNTVSNFQGPYTVTFQDNASFS  
GAIGYSVPSRLRLESGSEKFDVKDQYSAKDAFRFALARTSTTVTPDAQKYTM  
KNNGLSVASIMINGCYDLSENNAVSPYICAGIGEDFIEFDLHLIKLAYQKGLGISY  
YFPFKINVPAGYHHRVIGNKFNLNHNVTLUDEFPKATSAVATLNVATFGEAGVK  
FTP"  
8300..8890  
/codon\_start=1  
/transl\_table=11  
/product="unknown function UN3"  
/protein\_id="AAK28666.1"  
/db\_xref="GI:13511837"  
/translation="MIYKEKLRVGEYILAYLSFILSTYIFLVLNIRNYSLAICVI  
SLRITFNVTSTKLLKDKCRDTKFSNMCYLGKPLNLQIFYGIFSFIRNFOUNTLI  
IPNDSKCGFTVTLWDPALHYTVTLTGSEVRNFPDILYENIICQCKLLINYSRVLNQ  
HNKNTLVIIPIPNAREFSNIRVRNISINKESSEYEC"  
8880..9731  
/gene="omp-1w"  
8880..9731  
/gene="omp-1w"  
/note="omp-1 family member"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein OMP-1w"  
/protein\_id="AAK28667.1"

Best Local Similarity: 38.33% Mismatches: 98  
 Query Match: 31.02% Indels: 38  
 DB: 1 Gaps: 12

US-10-062-624-40 (1-293) x ECU72291 (1-27190)

Qy 1 MetAsnAsnLysLeuLysPheThrIleLeuValLeuValCysLeuLeuSer--- 19  
 Db 23451 ATGAATTCACAA---AAATTT---GTTAGGTGTGTGATGGCTACATCTCTTCTTC 23398  
 Qy 20 LeuProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyTyTyGlyLeu 39  
 Db 23397 TTGCTGATATCTTTTCTGATGCAACGTTCTCTGAAGGAGAAAG---GGATTT 23344  
 Qy 40 TyrlleSerGlyGlnTyLysProSerValSerValPheSerAsnPheSerValLysGlu 59  
 Db 23343 TATGTAGGTACTCAATATAAAGTTGGTGTCTTAATTTTCAGTAATTTTCAGTGAAGAA 23284  
 Qy 60 Thr---AsnValIleThrLysAsnLeuLeuAlaLeuLysLysAspValAspSerIleGlu 78  
 Db 23283 ACCTTCCTGAGCTCACAAAAGTATTTTCTTGTAGCCTGTGATAATCTAGC----- 23230  
 Qy 79 ThrLysThrAspAlaSerValGlyLysSerAsnProSerAsnPheThrIleProTyThr 98  
 Db 23229 -----ATTAGTGACCATGCGAGTTTTCACACAGCGTATAT 23194  
 Qy 99 AlaValPheGlnAspAsnSerValAsnPheAsnGlyThrIleGlyTyThrPheAlaGlu 118  
 Db 23193 CCTACATATGCAAGCAATTTTGTGGTGTGGTGTATTTGGATAT---TATGTTAAT 23137  
 Qy 119 GlyThrArgValGluLeuGlySerTyGluGluPheAspValLysAsn----- 135  
 Db 23136 GATTTAGTAGTAGTTTGAAGGTGCTTATGAAATTTTGAACCCGAGAGAGTGGTAT 23077  
 Qy 136 ProGlyGlyTyThrLeuSerAspAlaTyArgTyThrPheAlaLeuAlaArgGlu----- 153  
 Db 23076 CCTGAAGA-----GGAGAAAGCATAAGTTTGTCTATCTCTGTCGTGAGTCCACT 23026  
 Qy 154 MetLysGlyAsnSerPheThrProLysGlyLysValSerAsnSerIlePheHisThrVal 173  
 Db 23025 GTTCAAGATAATAAGTTT-----ATAGTG 23002  
 Qy 174 MetArgAsnAspGlyLeuSerIleLeuSerValIleValAsnValCysTyAspPheSer 193  
 Db 23001 TTGAAATGATGGGTGTATTGATGAATCTCTTAATGTGAACCTTTTGTATGATATTCCT 22942  
 Qy 194 LeuAsnAsnLeuSerIleSerProTyThrCysGlyGlyAlaGlyValAspAlaIleGlu 213  
 Db 22941 CATGTAGTATTCCTTTAGCACCATATATGTGTGCTGTGTGGAGCAGATTACATAAAA 22882  
 Qy 214 PhePheAspValLeuHisIleLysPheAlaTyThrGlnSerLysLeuGlyIleAlaTySer 233  
 Db 22881 TTTTGGGAATATCATTCCTCAAAATTTCTTATCAAGTGAATTTGGGGTTAACTATCCT 22822  
 Qy 234 LeuProSerAsnIleSerLeuPheAlaSerLeuTyThrHisLysValMetCysAsnGln 253  
 Db 22821 GTGAGTGTATGTTATGTTATTTGGTGGAGGTTATATCATAGGTTATAGGTAATAGG 22762  
 Qy 254 PheLysAsnLeuAsnVal---GlnHisValAlaGluLeuAlaSerIleProLysIleThr 272  
 Db 22761 TATGAACAGTAGAATAGCTTATCATCTGCCACATTAACGAATGTGCTCAAACTACT 22702  
 Qy 273 SerAlaValAlaThrLeuAsnIleGlyTyThrPheGlyGlyGluIleGlyAlaArgLeuThr 292  
 Db 22701 TCAGCATCTGCTACTTTAGATACTGATTATTTTGGTGGGAAGTTGGTATGAGATTTACA 22642  
 RESULT 22  
 AY028378  
 LOCUS AY028378 873 bp DNA linear BCT 06-NOV-2001  
 DEFINITION Cowdria ruminantium major antigenic protein MAPI gene, complete cds.  
 ACCESSION AY028378  
 VERSION AY028378.1 GI:13378319

KEYWORDS Ehrlichia ruminantium.  
 SOURCE Ehrlichia ruminantium  
 ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Anaplasmataceae; Ehrlichia.  
 REFERENCE 1 (bases 1 to 873)  
 AUTHORS Allsopp,M.T., Dorfling,C.M., Maillard,J.C., Bensaid,A., Haydon,D.T., van Heerden,H. and Allsopp,B.A.  
 TITLE Ehrlichia ruminantium major antigenic protein gene (map1) variants are not geographically constrained and show no evidence of having evolved under positive selection pressure  
 JOURNAL J. Clin. Microbiol. 39 (11), 4200-4203 (2001)  
 MEDLINE 21539003  
 PUBMED 11682561  
 REFERENCE 2 (bases 1 to 873)  
 AUTHORS Bensaid,A., Allsopp,M.T., Maillard,J.-C., Chantal,I. and Allsopp,B.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-MAR-2001) Elevage et Medecine Veterinaire, Centre de Cooperation Internationale en Recherche Agronomique pour le Developpement., Domaine du Clos, Petit Bourg 97170, Guadeloupe  
 FEATURES  
 source  
 1. .873  
 /organism="Ehrlichia ruminantium"  
 /strain="Vosloo"  
 /db\_xref="taxon:779"  
 /notes="obtained from cultured organisms"  
 CDS  
 1. .873  
 /notes="outer membrane protein"  
 /codon\_start=1  
 /transl\_table=11  
 /product="major antigenic protein MAP1"  
 /protein\_id="AAK18727.1"  
 /db\_xref="GI:13378320"  
 /translation="MNCKIFITSLISLVFLGVSFSDVIOEDSNPAGSVYISAKY MPTASHFGKMSIKEDSKSTQVFLKKDWDGVKVPYSENNTNSLFTFKDYSRYENN PFLGAGATGYSVMNGPRIEVEVSEYTVDVKNPGNYKNDAMHMCALDTCQASATNGAT LASSVMTKNSLNTNLSIMLACVDIMLDGMPVSPYVCAGITGLDVSVINATNPKLSYQ GKLGISYINSEASIFIGGHFHRVIGNEFKDIATLKIFTATNKVSTVANPFPASATLD VCHFGIEIGGVFV"  
 BASE COUNT 303 a 134 c 155 g 281 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 2.19e-30 Length: 873  
 Score: 461.50 Matches: 109  
 Percent Similarity: 57.95% Conservative: 66  
 Best Local Similarity: 36.09% Mismatches: 106  
 Query Match: 30.85% Indels: 21  
 DB: 1 Gaps: 10  
 US-10-062-624-40 (1-293) x AY028378 (1-873)  
 Qy 1 MetAsnAsnLysLeuLysPheThrIleLeuValLeuValCysLeuLeuSer--- 19  
 Db 1 ATGAATTCGAAGAAATTTT-----ATCACAAGTACACTATATCATTAGTGTCTTT 54  
 Qy 20 LeuProAsnIleSerSerLysLysAlaIleAsnAsnAsnAlaLysLysTyTyTyGlyLeu 39  
 Db 55 TTACCTGGTGTGCTCTTTCTGATGTAATACAGGAGGATAGACACCCAGCTGGTAGTGT 114  
 Qy 40 TyrlleSerGlyGlnTyLysProSerValSerValPheSerAsnPheSerValLysGlu 59  
 Db 115 TACATTAGCGCAAAATACATGCCCACTGCTTTCACACTTTGGTAAATATCATTAAGAA 174  
 Qy 60 ThrAsnValIleThrLysAsnLeuLeuAlaLeuLysLysAspValAspSerIleGluThr 79  
 Db 175 GATTCAAAAAGTACTCAACACAGTGTGTTGCTCTAAAAAAGATGCGGATGAGTAAAGTA 234  
 Qy 80 LysThrAspAlaSerValGlyIleSerAsnProSerAsnPheThr---IleProTyThr 98  
 Db 235 CCAACATCAGAAAAACACCAATTAATCTCT-----TCACCTTTTACTGAAAAAGATTATCT 288  
 Qy 99 AlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyThrPheAla 117



```

QY 219 HisIleLysPheAlaTyrglnSerLysLeuGlyIleAlaTygSerLeuProSerAsnIle 238
Db 1016 AATCTAAATATCTATACGAAGCTAGGCAATAGTACTCAATCAATCTGAAGCT 1075
QY 239 SerLeuPheAlaSerLeuTygHisLysValMetGlyAsnGlnPheLysAsnLeuAsn 258
Db 1076 TCTATCTTATCGGTGGACATTTCCATAGAGTATAGGTAAATTAATA----- 1126
QY 259 ValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSer----- 273
Db 1127 -----GATATTGCTACCTTAAAAAATATTACTTCAAAAACAGGAATATCT 1171
QY 274 -----AlaValAlaThrLeuAsnIleGlyTygPheGlyGlyGluIleGlyAla 289
Db 1172 AATCTGGCTTTGGCATCAGCAACACTTGATGTTGTCACTTTGGTATGAAATTTGGAGA 1231
QY 290 ArgLeuThrPhe 293
Db 1232 AGGTTTGATTT 1243

RESULT 24
AF368000 831 bp DNA linear BCT 06-NOV-2001
LOCUS Cowdria ruminantium isolate Blaaukrans major antigenic protein MAP1
DEFINITION gene, partial cds.
ACCESSION AF368000
VERSION AF368000.1 GI:15429315
KEYWORDS
SOURCE
ORGANISM
Ehrlichia ruminantium.
Ehrlichia ruminantium.
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
1 (bases 1 to 831)
Allsopp,M.T., Dorfling,C.M., Maillard,J.C., Bensaid,A.,
Haydon,D.T., van Heerden,H. and Allsopp,B.A.
Ehrlichia ruminantium major antigenic protein gene (map1) variants
are not geographically constrained and show no evidence of having
evolved under positive selection pressures
J. Clin. Microbiol. 39 (11), 4200-4203 (2001)
21539003
11682561
2 (bases 1 to 831)
Allsopp,M.T., Dorfling,C., Maillard,J.-C., Bensaid,A., Van
Heerden,H. and Allsopp,B.A.
Direct Submission
Submitted (05-APR-2001) Molecular Biology, Onderstepoort Veterinary
Institute, Old Soutpansberg Road, Onderstepoort, Gauteng 0110,
South Africa

FEATURES
Location/Qualifiers
1..831
/oranism="Ehrlichia ruminantium"
/isolate="Blaaukrans"
/db_xref="taxon:779"
<1..>831
/codon_start=1
/transl_table=1
/product="major antigenic protein MAP1"
/protein_id="AAK98142.1"
/db_xref="GI:15429316"
CDS
/transl_table=1
/product="major antigenic protein MAP1"
/protein_id="AAK98142.1"
/db_xref="GI:15429316"
GMSIKEDSKNTQTVFGLKKDWDGVKVPSTSENNSNTIFTEKDYSPRYENNPFGLFAG
AIGYNNGPRIEFESYETFDVKNPGNYKNDAMCYCALDTAOPTSNOGATLASSVMV
KNEUTDISLMNACVDIMLDGMPVSPYVCAGIGTDLVSVINATPKLSYQKGLGISY
SINPESAFIFGGHFRHVRIGNFKDITSTSKASSIPNPGFASATLDDVCHFGIE
I"

BASE COUNT 292 a 137 c 141 g 261 t
ORIGIN
Alignment Scores:
Pred. No.: 3,05e-30 Length: 831
Score: 459.50 Matches: 104
Percent Similarity: 57.09% Conservative: 61
Best Local Similarity: 35.99% Mismatches: 101

```

```

Query Match: 30.72% Indels: 23
DB: 1 Gaps: 8
US-10-062-624-40 (1-293) x AF368000 (1-831)
QY 10 IleAsnThrValLeuValCysLeuLeuSer---LeuProAsnIleSerSerLysAla 28
Db 1 ATCAAGTACACATATATCATATAGTGTCTATTTTACCTGGTGTCTTCTTCTGATGTA 60
QY 29 IleAsnAsnAlaLysLysTygTygLeuTygIleSerGlyGlnTygLeuProSer 48
Db 61 ATACAGGAAGACAGCAGCCAGCAGCGAGTGTTCATATTAGCGCAAAATACATGCCAAT 120
QY 49 ValSerValPheSerAsnPheSerValLysGluThrAsnValIleThrLysAsnLeu 68
Db 121 GCATCACATTTGTGTAATGTCATCAAGAGGATTCAAAATAATCTCAACACAGTATTT 180
QY 69 AlaLeuLysLysAspValAspSerIleGluThrLysThrAspAlaSerValGlyIleSer 88
Db 181 GGTCTAAAAAAGATTGGGATGGGTTAAAGTACCAATCA-----GAA 225
QY 89 AsnProSerAsnPheThrIle-----ProTygThrAlaValPheGlnAspAsn 104
Db 226 AACAGTAACAATTTCTCAAAATTTTACTCAAAAGATTTATTTCTCAGATATGAAAAACAAT 285
QY 105 Ser---ValAsnPheAsnGlyThrIleGlyTygThrPheAlaGluGlyThrArgValGlu 123
Db 286 CCGTTTTTAGGTTTTGCTGGAGCAATGGATCTCAATG---AATGGCCCAAGAAATAGAG 342
QY 124 IleGluGlySerTygGluGluPheAspValLysAsnProGlyGlyTygThrLeuSerAsp 143
Db 343 TTTGAGTATCTCTATGAACTTTTGATGTCAAAATCCAGGTGGCACTATAAAATGAT 402
QY 144 AlaTygArgTygPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGlu 163
Db 403 GCACACATGTACTGTCTTTA-----GATACAGCAGCACGCAACTAGT 444
QY 164 LysValSerAsnSerIlePheHisThrValMet---ArgAsnAspGlyLeuSerIleIle 182
Db 445 AATCAAGGTGCAACATAGCTTCATCTGTATGTTAAATAAGAAATTTTAAACAGATATA 504
QY 183 SerValIleValAsnValCysTygAspPheSerLeuAsnLeuSerIleSerProTyg 202
Db 505 TCATTAAATGTTAAATGCATGTTATGATATAATGCTTGATGAATGCCAGTTTCTCCATAT 564
QY 203 IleCysGlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIleLysPhe 222
Db 565 GTATGTGCAGGTATTGGTACTGATTAGTGTAGTCAATTAATGCTACAAATCTTAAATA 624
QY 223 AlaTygGlnSerLysLeuGlyIleAlaTygSerLeuProSerAsnIleSerLeuPheAla 242
Db 625 TCTTATCAAGGAAGCTAGGCAATAGTACTCAATCAATCTCCTGAGCTTCTATCTTATT 684
QY 243 SerLeuTygTygHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGlnHisVal 262
Db 685 GTTGACATTTCCATAGAGTATAGGTAATCAATTTAAAGATATTGCTACCTCAAAATA 744
QY 263 -----AlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeu 278
Db 745 TTTACTTCAACTAGTAAAGCATCATCTATATCTATCTGCTTGTGATCAGCAACACTT 804
QY 279 AsnIleGlyTygPheGlyGlyGluIle 287
Db 805 GATGTCCTCCCAATTTCCGGTATAGAAAT 831

RESULT 25
AX042305
LOCUS AX042305 864 bp DNA linear PAT 23-NOV-2000
DEFINITION Sequence 1 from Patent WO0065063.
ACCESSION AX042305
VERSION AX042305.1 GI:11340965
KEYWORDS heartwater rickettsia.
SOURCE

```



```

ORGANISM Ehrlichia ruminantium
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE
AUTHORS Barbet,A.F., Bowie,M.V., Ganta,R.R., Burridge,M.J., Mahan,S.M.,
McGuire,T.C., Rurangirwa,P.R., Moreland,A.L., Simbi,B.H.,
Whitmire,W. and Alleman,A.R.
TITLE Nucleic acid vaccines against rickettsial diseases and methods of
use
JOURNAL Patent: WO 0065063-A 1 02-NOV-2000;
UNIVERSITY OF FLORIDA (US)
FEATURES
source Location/Qualifiers
1..864
/organism="Ehrlichia ruminantium"
/db_xref="taxon:779"
CDS
1..864
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC17072.1"
/db_xref="GI:11340966"
/translation="MNCKKIFITSLISVFLPGVFSDFVIOEDSNPAGSVIYSKY
MPTASHFGKMSIKEDSKNTQVFLGKKDWDGVTSPSSNTNSTIFTEKDYSPVENY
PFLCPAGALGYMNGPRIEFVSYETFDVKNLGGYKKNHMYCALDTRAGNSTNGAG
LTSVMKNENLTNLSMLNACYDMLDGPVSPVCAGIGTDLVSVINATNPKLSYO
GKLGISVINSSEASIFIGGPHRVIGNEFKDIATLKIFTSKTGISNPGFASATLDVCH
FGIEIGGRFVP"
BASE COUNT 307 a 139 c 149 g 269 t
ORIGIN
Alignment Scores:
Pred. No.: 3.51e-30 Length: 864
Score: 459.00 Matches: 110
Percent Similarity: 57.38% Conservative: 65
Best Local Similarity: 36.07% Mismatches: 100
Query Match: 30.68% Indels: 30
DB: 6 Gaps: 10
US-10-062-624-40 (1-293) x AX042305 (1-864)
QY 1 MetAsnLeuLysLeuLysPheThrIleAsnThrValLeuValCysLeuLeuSer--- 19
Db 1 ATGAATTGCAAGAAATTTT-----ATCAAGTACACTAAATATCATAGTGCTATT 54
QY 20 LeuProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyTyTyGlyLeu 39
Db 55 TTACTCTGGTGCTCTTTCTGATGTAATACAGGACACACCCAGCAGGAGTGT 114
QY 40 TyrlSerGlyGlnTyTyLysProSerValSerValPheSerAsnPheSerValLysGlu 59
Db 115 TACATTAGCGCAAAATACATCGCAACTGCATCACATTTTGGTAAATGTCAATCAAAGAA 174
QY 60 ThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThr 79
Db 175 GATTCAAAAATACTCAACCGGTATTGGTCTAAAGAAAGATGGGATGGCGTTAAACAA 234
QY 80 LysThrAspAlaSerValGlyIleSerAsnProSerAsnPheThr---IleProTyTyThr 98
Db 235 CCATCAGATTCTAGC-----AATACTAATTTCTACATTTTCTGAAAAAGACTATTCT 288
QY 99 AlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyTyThrPheAla 117
Db 289 TTCAGATAGAAAACAATCCGTTTTTAGGTTTCGTCGGAGCAATGGGTACTCAATG--- 345
QY 118 GluGlyThrArgValGluIleGluGlySerTyTyGluGluPheAspValLysAsnProGly 137
Db 346 AATGGACCAAGATAGAGTTCCGAAATGATCCATGAACCTTTTGATGTAATAAAACCTAGGT 405
QY 138 GlyTyTyThrLeuSerAspAlaTyArgTyTyPheAlaLeuAlaArgGluMetLysGlyAsn 157
Db 406 GGCAACTATAAAACAACAGCACACATGCTACTGTGCTTTA-----GATACAGCAGCACAA 459
QY 158 SerPheThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAsp 177

```

```

Db 460 AATAGCACTAATGCGCAGGATTAACTACATCTGTT-----ATGGTAAAAACGAA 510
QY 178 GlyLeuSerIleIleLeuValAsnValCysTyTyAspPheSerLeuAsnLeu 197
Db 511 AATTTACAAATATATCATTAATGCGGTGTTATGATCATGCTTGTATGGGAATA 570
QY 198 SerIleSerProTyTyIleCysGlyGlyAlaGlyValAspAlaIleGluPheAspVal 217
Db 571 CCAGTTTCTCATATGATGTCAGGTATTGGCACTGACTTAGTGTAGTAAATTAATGCT 630
QY 218 LeuHisIleLysPheAlaTyTyGlnSerLysLeuGlyIleAlaTyTySerLeuProSerAsn 237
Db 631 ACAATCTCTAAATATCTTATCAAGGAAGCTAGGCATAAGTTACTCAATCAATCTGAA 690
QY 238 IleSerLeuPheAlaSerLeuTyTyTyHisLysValMetGlyAsnGlnPheLysAsnLeu 257
Db 691 GCTTCTATCTTATCGGTGGACATTTCCATAGATTTATAGTAAATGATTAATTA--- 744
QY 258 AsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSer----- 273
Db 745 -----GATATTGCTACCTTAAATAATATTACTTCAAAAACAGGAATA 786
QY 274 -----AlaValAlaThrLeuAsnIleGlyTyTyPheGlyGlyGluIleGly 288
Db 787 TCTAATCCTGGCTTTGTCATCAGCAACACTTGATGTTTGTCTACTTGTGTATAGAAATGGA 846
QY 289 AlaArgLeuThrPhe 293
Db 847 GGAAGGTTTGTATT 861
RESULT 26
AF368012 840 bp DNA linear BCT 06-NOV-2001
LOCUS Cowdria ruminantium isolate Omatjenne major antigenic protein MAP1
DEFINITION Gene, partial cds.
ACCESSION AF368012
VERSION AF368012.1 GI:15429335
KEYWORDS
SOURCE
ORGANISM Ehrlichia ruminantium.
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE
AUTHORS Allsopp,M.T., Dorfling,C.M., Maillard,J.C., Bensaïd,A.,
Haydon,D.T., van Heerden,H. and Allsopp,B.A.
TITLE Ehrlichia ruminantium major antigenic protein gene (map1) variants
are not geographically constrained and show no evidence of having
evolved under positive selection pressure
J. Clin. Microbiol. 39 (11), 4200-4203 (2001)
JOURNAL MEDLINE 21539003
PUBMED 11682561
REFERENCE
AUTHORS Allsopp,M.T., Dorfling,C., Maillard,J.-C., Bensaïd,A., Van
Heerden,H. and Allsopp,B.A.
TITLE Direct Submission
JOURNAL Submitted (05-APR-2001) Molecular Biology, Onderstepoort Veterinary
Institute, Old Sourpansberg Road, Onderstepoort, Gauteng 0110,
South Africa
FEATURES
source Location/Qualifiers
1..840
/organism="Ehrlichia ruminantium"
/isolate="Omatjenne"
/db_xref="taxon:779"
<1..5840
/codon_start=1
/transl_table=11
/product="major antigenic protein MAP1"
/protein_id="AAK98152.1"
/db_xref="GI:15429336"
/translation="FITSALISLVSFLPGVFSDFVIOEDITQKASVVISAKVYPTASH
FGKMSIKEDSRDTKTVFGIKDWDGKVPDTSSESNNSNNSIFTEKDYSPFKNPFLG
FAGAIGVMNGPRIEFVSYETFDVKNPGDKNDHMYCALDITQOSTYSTYGATLASS

```

VMVKNENLTDLSMLNACYDITLEGIPVSYICAGICTDLVSVINATNPKLSYQKLG  
ISYINPESAFIGHFRVIGNEFKDITTSKIFTSKASSIFSPGPASAILDVCHF  
GIEI

BASE COUNT 292 a 138 c 143 g 267 t  
ORIGIN

Alignment Scores:  
Pred. No.: 5,538-30 Length: 840  
Score: 456.50 Matches: 103  
Percent Similarity: 58.2% Conservative: 63  
Best Local Similarity: 36.14% Mismatches: 106  
Query Match: 30.51% Indels: 13  
DB: 1 Gaps: 7

US-10-062-624-40 (1-293) x AF368012 (1-840)

Qy 10 lleAsnThrValLeuValCysLeuLeuSer---LeuProAsnIleSerSerIlysAla 28  
Db 4 ATAACAAGTGCACATATATATAGTCTTCTTACCTGGAGTATCTTTCTGATCCA 63  
Qy 29 lleAsnAsnAlaLysLysTyrTyrGlyLeuTyrIleSerGlyGlnTyrIysProSer 48  
Db 64 GTACAGGAGGATATTACGCAAAAGCTAGCGTTTACATTAGTGCACAGTACGTGCCAACT 123  
Qy 49 ValSerValPheSerAsnPheSerValIysGluThrAsnValIleThrIysAsnLeuIle 68  
Db 124 GCATCACATTTTGGTAAATGTCATCAAGAAGATCTCCGAGATACATAACAGTATTT 183  
Qy 69 AlaLeuLysLysAspValAspSerIleGluThrLysThrAspAlaSerValGlyIleSer 88  
Db 184 GGTATATAAAGATGGTGGAGTAAAGTCCCAACATCAGAAAGTAAACAGTATAAC 243  
Qy 89 AsnProSerAsnPheThr---lleProTyrThrAlaValPheGlnAspAsnSer---Val 106  
Db 244 AATCTCTTATTTTACTGAAAAGACTATTCTTTTAAATATGAAAACAATCCATTTTA 303  
Qy 107 AsnPheAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArgValGluIleGluGly 126  
Db 304 GGTTTTGTGGAGCAATTGGATCTCAATG---AATGGCCCAAGAAATAGAGTTCGAAGTA 360  
Qy 127 SerTyrGluGluPheAspValIysAsnProGlyGlyTyrThrLeuSerAspAlaTyrArg 146  
Db 361 TCTTATGAACCTTCGATGTAATAAATCCAGGTGATACATAATAAAGTATGCACATATG 420  
Qy 147 TyrPheAlaLeuAlaArgGluMetLysGlyAsnSerPheThrProLysGluLysValSer 166  
Db 421 TACTGTGCCTTA-----GATACAACAACACAGAGTAGTACTTATGGTGCACATTAGCT 474  
Qy 167 AsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSerValIleVal 186  
Db 475 TCATCTGTIT-----ATGTAATAAATGAATAATTTAACAGACATTTTCAATAAGTTA 525  
Qy 187 AsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGly 206  
Db 526 AACGCATGCTATGATATACCTTGAAGGGATACCGGTTTCTCCATACATATGTCAGGT 585  
Qy 207 AlaGlyValAspAlaIleGluPhePheAspValLeuHisIleLysPheAlaTyrGlnSer 226  
Db 586 ATTGGCAGCTGATTAGTATCAGTAATCAATCAATGCTACAAATCTTAAACTATCTTATCAAGGA 645  
Qy 227 LysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyr 246  
Db 646 AAGTTAGGTATTAGTTATTCAATCAATCTCGAAGCTTCTATCTTATAGTGGGCATTTT 705  
Qy 247 HisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGlnHisVal-----262  
Db 706 CATAGGTTATAGGTAATCAATTAAGACATTACTACTTCTTAAGATATTACCTCAACT 765  
Qy 263 AlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIleGlyTyr 282  
Db 766 AGTAAAGCATCCAGTATACATAGTCCAGGCTTTGTCAGGATCAGCAATACCTGATGTTGCCAC 825  
Qy 283 PheGlyGlyGluIle 287

Db 826 TTCGGTATTGAATT 840

RESULT 27  
LOCUS

CRMAP1 1467 bp DNA linear BCT 10-OCT-1994  
DEFINITION C.ruminantium map1 gene.  
ACCESSION X74250  
VERSION X74250.1 GI:454266  
KEYWORDS immunodominant protein; major antigen; map1 gene; outer membrane protein.  
SOURCE heartwater rickettsia.  
ORGANISM Ehrlichia ruminantium  
Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
Anaplasmataceae; Ehrlichia.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE

JOURNAL  
MEDLINE  
PUBMED  
FEATURES

Location/Qualifiers

1. 1467  
/organism="Ehrlichia ruminantium"  
/isolate="Senegal stock"  
/db\_xref="taxon:779"  
481. 1335  
/gene="map1"  
481. 1335  
/gene="map1"  
/codon\_start=1  
/transl\_table=1  
/product="major antigenic protein"  
/protein\_id="CAA52309.1"  
/db\_xref="GI:454267"  
/db\_xref="SPTREMBL:O46327"  
/translation="MNCKKIFITSLISLVSFLPGVSFSDVIQENNPVGSVVISAKY  
MPTASHFGKMSIKEDSRDTRKAVFGKDKDGVKTPSGNTSIFTEKDIYSKIENPFLL  
GFAGAVGYSNNGPRIEFSEYFVDVNPNGNKNDAHYCALDNTASSSTAGATYSVM  
VKENLTDLSMLNACYDITLEGIPVSYICAGICTDLVSVINATNPKLSYQKLGIS  
YSINPEASIFIGHFRVIGNEFKDITTSKIFTSKASSIFSPGPASAILDVCHF  
EIGGRFVF"

BASE COUNT 493 a 204 c 242 g 528 t  
ORIGIN

Alignment Scores:

Pred. No.: 1,07e-29 Length: 1467  
Score: 456.50 Matches: 105  
Percent Similarity: 54.15% Conservative: 58  
Best Local Similarity: 34.88% Mismatches: 113  
Query Match: 30.51% Indels: 25  
DB: 1 Gaps: 8

US-10-062-624-40 (1-293) x CRMAP1 (1-1467)

Qy 1 MetAsnLysLeuLysPheThrIleIleAsnThrValLeuValCysLeuLeuSer--- 19  
Db 481 ATGAATTGCAAGAAATTTT-----ATAACAAGTACACATAATATCATTTAGTGTATT 534  
Qy 20 LeuProAsnIleSerSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeu 39  
Db 535 TTACTGTGTGTGATTTTCTGATGTAATCAGGAAGAACCAATCCAGTAGTAGTGT 594  
Qy 40 TyrIleSerGlyGlnTyrLysProSerValSerValPheSerAsnPheSerValLysGlu 59

Db 595 TATATTAGCGCAAAATACATGCACTGCAATCATCTTTTGGTAAATGTCAATTAAGAA 654

Qy 60 ThrAsnValIleThrLysAsnLeuLeuAlaLeuLysLysAspValAspSerIleGluThr 79

Db 655 GATTTCTAGATACATAAAGCAGTATTTGGCTTTAAAAAAGATTGGATGGAGTTAAACA 714

Qy 80 LysThrAspAlaSerValGlyIleSerAsnProSerAsnPheThrIleProTyrThrAla 99

Db 715 CATCGGGTAACCAATTCATTTTACTGAAAAAGACTATCTCTTCAATAT----- 768

Qy 100 ValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyrThrPheAlaGlu 118

Db 769 -----GAAATAATCCATTTTAGTGTTCAGAGCAGTTCGATATCTCAATG---AAT 819

Qy 119 GlyThrArgValGluIleGlySerTyrGluGluPheAspValLysAsnProGlyGly 138

Db 820 GGACCAAGAAATAGATTGAACTATCTTATGAACCTTTTCGACGTAAAGAAATCCAGCGGT 879

Qy 139 TyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSer 158

Db 880 AACTACAAAAACGATGCATATGTTGTCTAGAT----- 918

Qy 159 PheThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGly 178

Db 919 ---ACAGCATCAAGCTCTACTGCAGGCACTACATCTGTATGTTAAAAAATGAAAT 975

Qy 179 LeuSerIleIleSerValIleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSer 198

Db 976 TTAACAGATATTTTCATTAATGCTAAACGATGTTATGATCAATAATGCTTGACCGAATGCCA 1035

Qy 199 IleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIleGluPheAspValLeu 218

Db 1036 GTTCTCCATATGATGTGCGAGCATCTGCTAGCTTAGTATCATAGTAATTAACGCTACA 1095

Qy 219 HisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIle 238

Db 1086 AATCTTAATTAATCTTATCAAGGAATTAAGTAATGATTAATCAATAAACCCGGAAGCT 1155

Qy 239 SerLeuPheAlaSerLeuTyrThrHisLysValMetGlyAsnGlnPheLysAsnLeuAsn 258

Db 1156 TCTATCTTTTATGGTGGCAGCTTCCATAGATCATAGTAAAGTAATTAAGATATTGCT 1215

Qy 259 ValGlnHisVal-----AlaGluLeuAlaSerIleProLysIleThr 272

Db 1216 ACTTCTAAAGTTTTTACTAGCAGTGCTAATGCCAGTAGTGTGTAGTCCAGGTTTGCA 1275

Qy 273 SerAlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGluIleGlyAlaArgLeuThr 292

Db 1276 TCAGCAATA-----CTTGATGTTTGGCCACTTCGGCATAGAAATGGGAGGAGTTTGTA 1329

Qy 293 Phe 293

Db 1330 TTT 1332

RESULT 28

CRU50831

LOCUS

DEFINITION

Cowdria ruminantium Crystal Springs major antigenic protein 1

ACCSSION

U50831

VERSION

U50831.1 GI:1418251

KEYWORDS

heartwater rickettsia strain=Crystal Springs.

SOURCE

Ehrlichia ruminantium

ORGANISM

Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

ANAPLASMATACEAE; Ehrlichia.

REFERENCE

1 (bases 1 to 1263)

Reddy,G.R., Sulsona,C.R., Harrison,R.H., Mahan,S.M., Burridge,M.J.

and Barbet,A.F.

Sequence heterogeneity of the major antigenic protein 1 genes from

Cowdria ruminantium isolates from different geographical areas

Clin. Diagn. Lab. Immunol. 3 (4), 417-422 (1996)

96400830

PUBMED 8807206

2 (bases 1 to 1263)

Reddy,G.R.

Direct Submission

Submitted (06-MAR-1996) Roman G. Reddy, Pathobiology, University of

Florida, 471 Mowry Rd., Gainesville, FL 32611-0880, USA

Location/Qualifiers

1. 1263

/organism="Ehrlichia ruminantium"

/strain="Crystal Springs"

/db\_xref="taxon:779"

381..1244

/gene="map1"

381..1244

/gene="map1"

/note="surface protein"

/codon\_start=1

/transl\_table=1

/product="major antigenic protein 1"

/protein\_id="AAC44142.1"

/db\_xref="GI:1418252"

/translation="MNCKKIFITSLISVFLPGVSFSDVIOEDSNPAGSVIISAKY

MPTTSHFGKMSIKEDSKNTQTVGLKXDMGVKTPSSDNTNSITFTEKDYSPRYENN

PFSGAIGTSMNGPGIEFEVSTETFDVKNPNNYKNNAHMYCALDTAAGNSTNGAG

LTTSVMVKNLNLNISMLNACYDILMDGIPVSPYVCAGITGDLVINATNPKLSYQ

GLGISYINSEASIFIGGPHRVIGNEFKDIATLKIFTSKTGISNPGFASATLDVCH

FGIEIGRFFVS"

BASE COUNT 440 a 178 c 193 g 452 t

ORIGIN

Alignment Scores:

Pred. No.: 9.86e-30 Length: 1263

Score: 456.00 Matches: 107

Percent Similarity: 57.14% Conservative: 65

Best Local Similarity: 35.55% Mismatches: 101

Query Match: 30.48% Indels: 28

DB: 1 Gaps: 9

US-10-062-624-40 (1-293) x CRU50831 (1-1263)

Qy 2 AsnAsnLysLeuLysPheThrIleAsnThrValLeuValCysLeuLeuSer---Leu 20

Db 378 AATATGAATTGCAAGAAAAATTTTATCAAGTACACTAATATCATGTGTCATTTTA 437

Qy 21 ProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyrTyrGlyLeuTyr 40

Db 438 CCTGGTGTCTCTTTCTGATGTAATACAGAGACAGCAACCCAGGAGGAGTTTAC 497

Qy 41 IleSerGlyIntYrLysProSerValSerValPheSerAsnPheSerValLysGluThr 60

Db 498 ATTAGCGCAAAATACATGCCAATCATCATTTTGGTAAATGTCAATCAAGAAGAT 557

Qy 61 AsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThrLys 80

Db 558 TCRAAAAAATACTCAACAGATTTTGGTCTAAAAAAGATTGGGATGGCGTTAAACACCA 617

Qy 81 ThrAspAlaSerValGlyIleSerAsnProSerAsnPheThr---IleProTyrThrAla 99

Db 618 TCAGATTTCTAGC-----AATACTAATTTCTCAATTTTACTGAAAAAGACTATCTTTC 671

Qy 100 ValPheGlnAspAsnSerValAsn---PheAsnGlyThrIleGlyTyrThrPheAlaGlu 118

Db 672 AGATATGAAAAACAATCCGTTTTCAGGTTTCGCTGGAGCAATTTGGTACTCAATG---AAT 728

Qy 119 GlyThrArgValGluIleGlyLysTyrGluGluPheAspValLysAsnProGlyGly 138

Db 729 GGACCAAGAAATAGATTTCGAAGTATCTATGAAACTTTTGTATGTAAAAAACCAGGTCGC 788

Qy 139 TyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArgGluMetLysGlyAsnSer 158

Db 789 AACTATAAACAACACGACACATGTACTGTGCTTTA-----GATACAGCAGCACAANAAT 842

Qy 159 PheThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAspGly 178

```

Db      843  ACCATAATGCCGAGGATTAACATCTGTT-----ATGGTAAAAACGAAAT 893
Qy      179  LeuSerIleIleSerValIleValAenValCysTyrAspPheSerLeuAenLeuSer 198
Db      894  TTAACAATATATCATTAATGTTAAATGCGTGTATGATATCATGCTTGATGGAATACCA 953
Qy      199  IleSerProTyrIleCysGlyAlaGlyValAlaIleGluPhePheAenValLeu 218
Db      954  GTTCTCCATATGATGTGCGAGTATGCGCACTTAGTGTAGTCAATTAATGCTACA 1013
Qy      219  HisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyrSerLeuProSerAenIle 238
Db      1014  AATCTTAATTAATTAATCAAGAAAGCTGAGCAATAGTACTCAATCAATCTGAAGCT 1073
Qy      239  SerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAenGlnPheLysAenLeuAen 258
Db      1074  TCTATCTTTATCGGTGGACATTTCCATAGAGTTATAGGTAAATGAATTTAAA----- 1124
Qy      259  ValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSer----- 273
Db      1125  -----GATATTCTCTACCTTAAATAATATTACTTCAAAACAGGAATATCT 1169
Qy      274  -----AlaValAlaThrLeuAenIleGlyTyrPheGlyGlyGluIleGlyAla 289
Db      1170  AATCTCGGCTTGCATGCACAACTTGATGTTGTCTACTTTGGTATAGAAATGGAGGA 1229
Qy      290  Arg 290
Db      1230  AGG 1232

RESULT 29
AF368009
LOCUS
DEFINITION
Cowdria ruminantium isolate Morgenswag1 major antigenic protein
ACCESSION
AF368009
VERSION
AF368009.1 GI:15429329
KEYWORDS
Ehrlichia ruminantium.
SOURCE
Ehrlichia ruminantium
ORGANISM
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE
1 (bases 1 to 840)
Allopp,M.T., Dorfling,C.M., Maillard,J.C., Bensaïd,A.,
Haydon,D.T., van Heerden,H. and Allsopp,B.A.
Ehrlichia ruminantium major antigenic protein gene (map1) variants
are not geographically constrained and show no evidence of having
evolved under positive selection pressure
J. Clin. Microbiol. 39 (11), 4200-4203 (2001)
21539003
MEDLINE
PUBMED
11682561
REFERENCE
2 (bases 1 to 840)
Allopp,M.T., Dorfling,C., Maillard,J.-C., Bensaïd,A., Van
Heerden,H. and Allsopp,B.A.
Direct Submission
Submitted (05-APR-2001) Molecular Biology, Onderstepoort Veterinary
Institute, Old Soutpansberg Road, Onderstepoort, Gauteng 0110,
South Africa
FEATURES
Location/Qualifiers
1..840
/organism="Ehrlichia ruminantium"
/isolate="Morgenswag1"
/species="goat"
/db_xref="taxon:779"
/note="detected in goat blood from the farm Morgenswag,
Northern Cape, South Africa"
<1..>840
/codon_start=1
/transl_table=11
/product="major antigenic protein MAP1"
/protein_id="AAK98149.1"
/db_xref="GI:15429330"
CDS

```

```

/translation="FITSALISLVSLFPGVSPDPVOEDITOKASVVISAKYVPTASH
FGKMSIKEDSRDTKTVFGIKDWDGVKVPFSESNNSNNSIFTEKDYSPKYENNPFLG
FAGAIGYSMMGPRIEFYSYETFDVKNPKNYKNDAMHYKALDITQOSSTYGLASS
VMVENELTDISLMNACVDTILEGIPYICAGITGDLVSNINATPKLSYQKLG
ISVSNPEASIFIGFHFVRVIGNEFKDITTSKASTSPSPGFASAILDVCHF
GIEI"
BASE COUNT      291 a   137 c   145 g   267 t
ORIGIN
Alignment Scores:
Pred. No.:      8.17e-30      Length:      840
Score:          454.50      Matches:    103
Percent Similarity: 57.89%      Conservative: 62
Best Local Similarity: 36.14%      Mismatches: 107
Query Match:    30.38%      Indels:    13
DB:             1          Gaps:      7
US-10-062-624-40. (1-293) x AF368009 (1-840)
Qy      10  IleAsnThrValLeuValCysLeuLeuSer---LeuProAenIleSerSerSerLysAla 28
Db      4  ATAAACAAGTCACATAATATCATTTAGTGTCTTCTTACCTGGAGTATCTTTTCTGATCCA 63
Qy      29  IleAenAenAenAlaLysLysTyrTyrGlyLeuTyrIleSerGlyGlnTyrLysProSer 48
Db      64  GTACAGAGGATATTACGCAAAAGCTAGGCTTACATTTAGTAGCAAGTAGTGCCTCAACT 123
Qy      49  ValSerValPheSerAenPheSerValLysGluThrAenValIleThrLysAenLeuIle 68
Db      124  GCATCACATTTTGGTAAATGTCAATCAAGAAGATTCCCGAGATACTAAACACAGATTTT 183
Qy      69  AlaLeuLysLysAspValAspSerIleGluThrLysThrAspAlaSerValGlyLysSer 88
Db      184  GGTATAAAAAAGATTGGGATGGAGTTAAAGTGCACACATCGAGAAAGTAACAGTAATAC 243
Qy      89  AsnProSerAenPheThr---IleProTyrTyrAlaValPheGlnAspAenSer---Val 106
Db      244  AATTCTCTATTTTACTGAAAAGACTATCTTTTAAATATGAAAACATCCATTTTTA 303
Qy      107  AsnPheAenGlyThrIleGlyTyrPheAlaGluGlyThrArgValGluIleGluGly 126
Db      304  GGTTTTGTGGAGCAATTGGATACTCAATG---AATGGCCCAAGATAGAGTTCGAAGTA 360
Qy      127  SerTyrGluGluPheAspValLysAenProGlyTyrThrLeuSerAenAlaTyrArg 146
Db      361  TCTTATGAACTTTTCGATGTAAAAAATCCAGGTGATACTATAAAAATGATGCACATG 420
Qy      147  TyrPheAlaLeuAlaArgGluMetLysGlyAenSerPheThrProLysGluLysValSer 166
Db      421  TACTGTGCCTTA-----GATACACACACAGAGTAGTAGTCTTATGGTGCAACATTAGCT 474
Qy      167  AsnSerIlePheHisThrValMetArgAenAspGlyLeuSerIleLeuSerValIleVal 186
Db      475  TCATCTGTT-----ATGGTAGAAAATGAAAAATTTAACAGACATTTCAATTAATGTTA 525
Qy      187  AsnValCysTyrAspPheSerLeuAenAenLeuSerIleSerProTyrIleCysGlyGly 206
Db      526  AACCCATGCTATGATATAAACCCCTTGAAGGGGATACCGGTTCTCCATACATATGCGAGGT 585
Qy      207  AlaGlyValAspAlaIleGluPhePheAenValLeuHisIleLysPheAlaTyrGlnSer 226
Db      586  ATTGGGACTGATTAGTATCATGTAATCAATGCTACAAATCTTAACATCTTATTCNAGGA 645
Qy      227  LysLeuGlyIleAlaTyrSerLeuProSerAenIleSerLeuPheAlaSerLeuTyrTyr 246
Db      646  AAGTTAGGTATTAGTTATTCAATCAATCTCGAAGCTTCTATCTTTATAGGTGGCATTTT 705
Qy      247  HisLysValMetGlyAenGlnPheLysAenLeuAenValGlnHisVal----- 262
Db      706  CATAGGGTTATAGGTAATGAATTTAAAGACATTACTTCTTAAGATATTCACTCAACT 765
Qy      263  AlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAenIleGlyTyr 282

```

```

Db      766 AGTAAAGCATCCAGTATACCTAGTCCAGGCTTTGGCATCAGCAATACCTTGATGTTGGCCAC 825
Qy      283 PheGlyGlyGluLe 287
Db      826 TTCGGTATTGAATT 840

RESULT 30
CRU50830
LOCUS      1263 bp DNA linear BCT 14-JUL-1996
DEFINITION Cowdria ruminantium Antigua major antigenic protein 1 (map1) gene,
complete cds.
ACCESSION U50830
VERSION   U50830.1 GI:1418249
KEYWORDS  heartwater rickettsia strain=Antigua.
SOURCE    Ehrlichia ruminantium
ORGANISM  Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE 1 (bases 1 to 1263)
AUTHORS   Reddy,G.R., Sulsona,C.R., Harrison,R.H., Mahan,S.M., Burridge,M.J.
and Barbet,A.F.
TITLE     Sequence heterogeneity of the major antigenic protein 1 genes from
Cowdria ruminantium isolates from different geographical areas
JOURNAL   Clin. Diagn. Lab. Immunol. 3 (4), 417-422 (1996)
MEDLINE   96400830
PUBMED    8807206
REFERENCE 2 (bases 1 to 1263)
AUTHORS   Reddy,G.R.
DIRECT SUBMISSION
TITLE     Submitted (06-MAR-1996) Roman G. Reddy, Pathobiology, University of
Florida, 471 Mowry Rd., Gainesville, FL 32611-0880, USA
JOURNAL   Location/Qualifiers
FEATURES  source
           1..1263
           /organism="Ehrlichia ruminantium"
           /strain="Antigua"
           /db_xref="taxon:779"
           397..1242
           /gene="map1"
           /gene="map1"
           /note="surface protein"
           /codon_start=1
           /transl_table=11
           /product="major antigenic protein 1"
           /protein_id="AAC44141.1"
           /db_xref="GI:1418250"
           /translation="MNCKIPITSLVSLVFLPGVSFSDVIOEDSPVGSVVISAKY
MPTASHFKKMSIKEDSRDTKVFLKKDWDGVKTSNNTIFTEKDYSEKYNPFLGF
AGAIGYMNWGRPRIEFSEISYETFDVKNPGNKINDAHYKCALDITATSGKGAASVNV
KNEMLDISLMNACYDIMLDMPVSPVCAGIGTDLVSNINPKLSYQGLKGISY
SINPEASIFIGGHFRVIGNEFKDITTSKIPNTSNTGATPGFASAILDVCHFIEIG
GRFVF"

BASE COUNT 430 a 153 c 205 g 475 t
ORIGIN
Alignment Scores:
Pred. No.: 2,15e-29 Length: 1263
Score: 452.00 Matches: 105
Percent Similarity: 54.70% Conservative: 58
Best Local Similarity: 35.23% Mismatches: 113
Query Match: 30.21% Indels: 22
DB: 1 Gaps: 8

US-10-062-624-40 (1-293) x CRU50830 (1-1263)
Qy      1 MetAsnLeuLysLeuLysPheThrIleLeuValLeuValCysLeuLeuSer--- 19
Db      397 ATGAATGCAGAAATTTT-----ATAACAAGTACACTAATATCATGTCATTT 450
Qy      20 LeuProAsnIleSerSerLysAlaIleAsnAsnAlaLysLysTyThrGlyLeu 39
Db      451 TTACTGTGTGTCCTTTCTTGATGTAATACAGGAAGACAGCCAGTAGGTGTT 510

```

```

Qy      40 TyrIleSerGlyGlnTyLysProSerValSerValPheSerAsnPheSerValLysGlu 59
Db      511 TACATCAGTCGCAAAATACATGCCAACTGCTTCACTTTTGGTAAATGTCAATTAAGAA 570
Qy      60 ThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAspSerIleGluThr 79
Db      571 GATTCTAGACATCAAGTAGATTGTTGGTCTAAAGAAAGATGGGATGGATTAACA 630
Qy      80 LysThrAspAlaSerValGlyIleSerAsnProSerAsnPheThr---IleProTyThr 98
Db      631 AGTAGTAGCAATCAAT-----TTCCTGAAAGAGATTAATCT 669
Qy      99 AlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIleGlyTyThrPheAla 117
Db      670 TTTAAATATGAAATAATCCATCTCTAGGTTTTCGTCGAGCAATTCGATACATG--- 726
Qy      118 GluGlyThrArgValGluIleGlySerTyGluGluPheAspValLysAsnProGly 137
Db      727 AATGTCACAAGATAGAAATTTGAAATATCTCTAGAAACTTTTGATGTAAAGAAACCCAGGT 786
Qy      138 GlyTyThrLeuSerAspAlaTyArgTyPheAlaLeuAlaArgGluMetLysGlyAsn 157
Db      787 GGTAACTATAAAATGATGCCATATGATGTGTCCTTAGACACAGCAACATCATCTGGT 846
Qy      158 SerPheThrProLysGluLysValSerAsnSerIlePheHisThrValMetArgAsnAsp 177
Db      847 GGAGCAGCAGCAAGTACATCTGTT-----ATGTAAGAAATGAA 885
Qy      178 GlyLeuSerIleIleSerValIleValAsnValCysTyRAspPheSerLeuAsnAsnLeu 197
Db      886 AATTAAACAGATATATCAATTAATGTTAAATGCATGTTATGATATATATGCTTGATGGAATG 945
Qy      198 SerIleSerProTyIleCysGlyGlyAlaGlyValAspAlaIleGluPheAspVal 217
Db      946 CCAGTGTCTCCATATGATGTCGAGGTATGGTACCGATTAGTGTCAGTAATTAATCT 1005
Qy      218 LeuHisIleLysPheAlaTyRLeuGlnSerLysLeuGlyIleAlaTyRLeuProSerAsn 237
Db      1006 ACAATCCTAAATATCTCTATCAAGGAAACTTGGTATAAGTTATTTCGATAAATCCAGAA 1065
Qy      238 IleSerLeuPheAlaSerLeuTyRHisLysValMetGlyAsnGlnPheLysAsnLeu 257
Db      1066 GCATCTATATTTATTTGCGGACATTTTCATAGAGTTATAGTAGTATGATTCGTAAGATATT 1125
Qy      258 AsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThrSer-----AlaVal 275
Db      1126 ACTACTTCTAAGATTTTAAATAGTAACTAGTAACACTGGTGGTCCCACTCCAGGCTTGCATCA 1185
Qy      276 AlaThrLeuAsnIleGlyTyRPhGlyGlyGluIleGlyAlaArgLeuThrPhe 293
Db      1186 GCAATACTCATGCTCGCCATTTCCGTTATAGAAATTTGGAGGAAGGTTTGTATT 1239

RESULT 31
AF393395
LOCUS      1282 bp DNA linear BCT 27-MAR-2002
DEFINITION Ehrlichia chaffeensis strain V9 outer membrane protein p28 gene,
complete cds.
ACCESSION AF393395
VERSION   AF393395.1 GI:15991543
KEYWORDS  Ehrlichia chaffeensis.
SOURCE    Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
ORGANISM  Anaplasmataceae; Ehrlichia.
REFERENCE 1 (bases 1 to 1282)
AUTHORS   Long,S.W., Zhang,X.F., Qi,H., Standaert,S., Walker,D.H. and Yu,X.J.
TITLE     Antigenic variation of Ehrlichia chaffeensis resulting from
differential expression of the 28-kilodalton protein gene family
JOURNAL   Infect. Immun. 70 (4), 1824-1831 (2002)
MEDLINE   21893092
PUBMED    11895944
REFERENCE 2 (bases 1 to 1282)
AUTHORS   Yu,X.-J. and Zhang,X.-F.

```

Qy	189	CysTyrAspPheSerLeuAenAenSerLeuPheAAspValLeuHisLeuPheAlaTyrGlnSerLysLeu	208
Db	730	TGCTATGATGTAATAAGTGAAGAATAACCTTTTCTCTTACATATGTGCAGGTATTGGT	789
Qy	209	ValAspAlaIleGluPhePheAspValLeuHisLeuPheAlaTyrGlnSerLysLeu	228
Db	790	ACTGATTAGTATCCATGTTTGAAGTTACAAACCCCTAAATTTCTTACCAAGGAAAGTTA	849
Qy	229	GlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrHisLys	248
Db	850	GGTTTAAGCTACTCTATAAGCCCGAGAACTCTGCTCTTTCGCGGACGACATTTCCATAAG	909
Qy	249	ValMetClyAsnGlnPheLysAsnLeuAenValGlnHisValAlaGluLeuAlaSerIle	268
Db	910	GTAATGGAAACAAATTCAGAGATATTTCCT-----GCTATAATA	948
Qy	269	ProLysIleThrSer-----AlaValAlaThrLeuAenIleGlyTyr	282
Db	949	CTAGTGGATCAGTCTTACAGGAAATCACCCTGCAATAGTAACTAGATGTATGCCAC	1008
Qy	283	PheGlyClyGluIleGlyAlaArgLeuThrPhe	293
Db	1009	TTTGGTATAGAACTTGGAGGAAGATTGCTTTC	1041
RESULT 32			
AF077734		1312 bp DNA linear	BCT 20-JUN-2001
LOCUS		Ehrlichia chaffeensis strain Sapulpa outer membrane protein p28	
DEFINITION		precursor, gene, complete cds.	
ACCESSION	AF077734		
VERSION	AF077734.2	GI:14495546	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REMARK			
COMMENT			
FEATURES			
source			
CDS			

```

BASE COUNT      437 a   219 c   224 g   432 t
ORIGIN

Alignment Scores:
Pred. No.:      2,48e-29      Length:      1312
Score:          451.50      Matches:      113
Percent Similarity: 52.73%      Conservative: 51
Best Local Similarity: 36.33%      Mismatches: 94
Query Match:    30.18%      Indels:      53
DB:            1          Gaps:      12

US-10-062-624-40 (1-293) x AF077734 (1-1312)

Qy      1 MetAsnAsnLysLeuLysPheThrIleAsnThrValLeuValCysLeuLeu---Ser 19
Db      243 ATGAATTACAAAAGTTTTC-----ATAACAAGTGCATTGCATCTACTATCTTCT 296

Qy      20 LeuProAsnIleSerSer-----LysAlaIleAsnAsnAlaLysLys 35
Db      297 CTACTGGAGTATCATTTTCGACCAACAGGTAGTGTATTAAACGGCAAT----- 347

Qy      36 TyrTyrGlyLeuTyrlleSerGlyGlnTyrLysProSerValSerValPheSerAsnPhe 55
Db      348 -----TTCATATCAGTGGAAATATATGCAAGCGCTTCGCATTTTGGGGTGT 398

Qy      56 SerValLysGluThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValasp 75
Db      399 TCTGTAGGAAGAAAGAAATACACAGCTGGAGTATTGGATTGAACGAAGATTGGCAT 458

Qy      76 SerIle-----GluThrLysThrAspAlaSerValGlyIleSerAsnProSerAsn 92
Db      459 GGCAGTGCATACCTCACACTCATAGTAATACTCATTTACTGTTCATAAC----- 509

Qy      93 PheThrIleProTyrThrAlaValPheGlnAspAsnSer---ValAsnAsnGlyThr 111
Db      510 -----TACTCATTTAAATATGAATAATATCCATTTCTAGGTTTTCAGGAGCT 557

Qy      112 IleGlyTyrThrPheAlaGluGlyThrArgValGluIleGluGlySerTyrGluGluPhe 131
Db      558 ATTGGCTACTCAATG---GATGGTCCAGAAATAGAGCTTGAAGTATCTTATGAGACATT 614

Qy      132 AspValLysAsnProGlyGlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAla 151
Db      615 GATGTAAAAATCAAGGTAAACAATATAGAACCAAGAGCTCATAGATATCTGCTTTTATCC 674

Qy      152 ArgGlu-----MetLysGlyAsnSerPheThrProLysGluLysValSerAsnSer 168
Db      675 CGTCAGGCCACATAGCAGCAGATAATACATT----- 707

Qy      169 IlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSerValIleValAsnVal 188
Db      708 -----GTTTTCCTAAAAAATGAAGGCTACTTGACATATCATTTATGCTGAACGCA 758

Qy      189 CysTyrAspPheSerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyValagly 208
Db      759 TGCTATGATTAATAAGTAAGGAATACCTTTTCTCCTTACATATGTCAGGATTGGT 818

Qy      209 ValAspAlaIleGluPhePheAspValLeuHisIleLysPheAlaTyrGlnSerLysLeu 228
Db      819 ACTGATTAGTATCCATCTTTGAAGTTACAAACCTTAAATTTCTTACCAAGGAAGTTA 878

Qy      229 GlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLys 248
Db      879 GGTTAAGCTACTCTATAAGCCAGAACTTCTGCTTTGTGCGGGACATTTCCATAAG 938

Qy      249 ValMetGlyAsnGlnPheLysAsnLeuAsnValGlnHisValAlaGluLeuAlaSerIle 268
Db      939 GTAATGGAAACGAATTCAGAGATATTCCCT-----GCTATAATA 977

Qy      269 ProLysIleThrSer-----AlaValAlaThrLeuAsnIleGlyTyr 282
Db      978 CCTAGTGCATCAAGTCTTACAGGGAATCACCTCGAATAGTAACACTAGATTGTCAC 1037

```

---

```

Qy      283 PheGlyGlyGluIleGlyAlaArgLeuThrPhe 293
Db      1038 TTTCGTATAGAACTTGGAGGAAGATTGCTTTC 1070

RESULT 33
AF393389      1280 bp      DNA      linear      BCT 27-MAR-2002
LOCUS      Ehrlichia chaffeensis strain V2 outer membrane protein p28 gene,
DEFINITION      complete cds.
ACCESSION      AF393389
VERSION      AF393389.1 GI:15991531
KEYWORDS
SOURCE      Ehrlichia chaffeensis.
ORGANISM      Ehrlichia chaffeensis
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE      1 (bases 1 to 1280)
AUTHORS      Long, S.W., Zhang, X.F., Qi, H., Standeert, S., Walker, D.H. and Yu, X.J.
TITLE      Antigenic variation of Ehrlichia chaffeensis resulting from
            differential expression of the 28-kilodalton protein gene family
            Infect. Immun. 70 (4), 1824-1831 (2002)
JOURNAL      MEDLINE
PUBMED      21893092
REFERENCE      2 (bases 1 to 1280)
AUTHORS      Yu, X.-J., Zhang, X.-F. and Walker, D.H.
TITLE      Direct Submission
JOURNAL      Submitted (20-JUN-2001) Pathology, University of Texas Medical
            Branch, 301 Univ. Blvd., Galveston, TX 77555-0609, USA
FEATURES
            source
            1..1280
            /organism="Ehrlichia chaffeensis"
            /strain="V2"
            /db_xref="taxon:945"
            242..1069
            /note="p28-19; 28 kDa outer membrane protein"
            /codon_start=1
            /transl_table=11
            /product="outer membrane protein p28"
            /protein_id="AA112919.1"
            /db_xref="GI:15991532"
            /translation="MNYKVFTISALISLISLPGVSFSDPTGSGINGNFVISCYMP
            SASHFGVFSKEERTTAGVFLGKQDWGDAIPIHTQSTSTFTVSNVSKYNNFLGF
            AGAIGVMDGPRIEVSFETFDVKNQNNYKNEAHRYCALSRQATADNKFVLKNE
            GLADISPLNADVISEGIPFPGYICAGITDLSVMFVETNPVKISYQKGLSYSIS
            PETSVFVGHFHKVMGNEFRDIPAIIPSGSNLTGNHPAIVLTLDVCHFGIELGRFAP"
BASE COUNT      428 a   207 c   219 g   426 t
ORIGIN

Alignment Scores:
Pred. No.:      2,65e-29      Length:      1280
Score:          451.00      Matches:      112
Percent Similarity: 53.42%      Conservative: 52
Best Local Similarity: 36.48%      Mismatches: 97
Query Match:    30.15%      Indels:      46
DB:            1          Gaps:      12

US-10-062-624-40 (1-293) x AF393389 (1-1280)

Qy      1 MetAsnAsnLysLeuLysPheThrIleAsnThrValLeuValCysLeuLeu---Ser 19
Db      242 ATGAATTACAAAAGTTTTC-----ATAACAAGTGCATTGCATCTACTATCTTCT 295

Qy      20 LeuProAsnIleSerSer-----LysAlaIleAsnAsnAlaLysLys 35
Db      296 CTACTGGAGTATCATTTTCGACCAACAGGTAGTGTATTAAACGGCAAT----- 346

Qy      36 TyrTyrGlyLeuTyrlleSerGlyGlnTyrLysProSerValSerValPheSerAsnPhe 55
Db      347 -----TTCATATCAGTGGAAATATATGCAAGCGCTTCGCATTTTGGGGTGT 397

Qy      56 SerValLysGluThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValasp 75
Db      398 TCTGTAGGAAGAAAGAAATACACAGCTGGAGTATTGGATTGAACGAAGATTGGCAT 457

```

Qy 76 Serlle-----GluThrLysThrAspAlaSerValGlyLysSerAsnProSerAsn 92  
 Db 458 GGCAGTCAATACCTCACATCAGTACTAGTACTCCATTACTGTTTCAAC----- 508  
 Qy 93 PheThrIleProTyrThrAlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThr 111  
 Db 509 -----TATCATTTAAATATGAAATATCCATTCTAGGTTTTCAGAGCT 556  
 Qy 112 IleGlyTyrThrPheAlaGluGlyThrArgValGluIleGluGlySerTyrGluGluPhe 131  
 Db 557 ATTGCTACTCAATG---GATGGTCCAGAAATAGAGCTTGAAGTATCTTATGAGACATT 613  
 Qy 132 AspValLysAsnProGlyGlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAla 151  
 Db 614 GATGTTAAATCAAGAGTAAACATATAGAAAGGCTCATAGACTACTGCTTTATCC 673  
 Qy 152 ArgGlu-----MetLysGlyAsnSerPheThrProLysGluLysValSerAsnSerlle 169  
 Db 674 COTCAGGCCACATAGCAGATTAATATTT----- 703  
 Qy 170 PheHisThrValMetArgAsnAspGlyLeuSerIleIleSerValIleValAsnValCys 189  
 Db 704 -----GTTTCTTAAATAATCAAGGCTACTTGACATATCATTTATGCTGAACGATGC 757  
 Qy 190 TyrAspPheSerLeuAsnAsnLeuSerlleSerProTyrIleCysGlyGlyVal 209  
 Db 758 TATGATGTAATAAGTGAAGGATACCTTTTCTCTTACATATGTCAGGTTTGGTACT 817  
 Qy 210 AspAlaIleGluPhePheAspValLeuHisIleLysPheAlaTyrGlnSerLysLeuGly 229  
 Db 818 GATTAGTATCCATGTTTGAAGTTACAAACCTTAAATTTCTTACAGGAAGTAGGT 877  
 Qy 230 IleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLysVal 249  
 Db 878 TTAAGTACTCTATAAGCCAGAACTCTGCTTTGTCGGCGGACATTTCCATAAGGTA 937  
 Qy 250 MetGlyAsnGlnPheLysAsnLeuValGlnHisValAlaGluLeuAlaSerllePro 269  
 Db 938 ATGGGAACGAATTCAGATATTT-----CCTGCTATATATACCTAGTGATCA 985  
 Qy 270 LysIleThrSer-----AlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGlu 286  
 Db 986 ATCTTACAGGAATCACCTGCATAGTAACTAGATGATGCCACTTTCGATAGNA 1045  
 Qy 287 IleGlyAlaArgLeuThrPhe 293  
 Db 1046 CTTGGAGGAAGATTGCTTTC 1066

RESULT 34  
 AF077733  
 LOCUS  
 DEFINITION 1243 bp DNA linear BCT 20-JUN-2001  
 Ehrlichia chaffeensis strain Jax outer membrane protein p28  
 precursor, gene, complete cds.  
 ACCESSION AF077733  
 VERSION AF077733.2 GI:14495545  
 KEYWORDS  
 SOURCE Ehrlichia chaffeensis.  
 ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Anaplasmataceae; Ehrlichia.

REFERENCE  
 1 (bases 1 to 1243)  
 Yu, X.-J., McBride, J.W. and Walker, D.H.  
 Genetic diversity of the 28-kilodalton outer membrane protein gene  
 in human isolates of Ehrlichia chaffeensis  
 J. Clin. Microbiol. 37 (4), 1137-1143 (1999)  
 MEDLINE 99175287  
 PUBMED 10074538

REFERENCE  
 2 (bases 1 to 1243)  
 Yu, X.-J. and Walker, D.H.  
 Direct Submission  
 TITLE Submitted (14-JUL-1998) Pathology, The University of Texas Medical  
 JOURNAL Branch, 301 Univ. Blvd., Galveston, TX 77555, USA

3 (bases 1 to 1243)  
 Yu, X.-J. and Walker, D.H.  
 Direct Submission  
 TITLE Submitted (20-JUN-2001) Pathology, The University of Texas Medical  
 JOURNAL Branch, 301 Univ. Blvd., Galveston, TX 77555, USA  
 REMARK  
 COMMENT Sequence update by submitter  
 On Jun 20, 2001 this sequence version replaced gi:3414968.  
 FEATURES  
 Location/Qualifiers  
 1..1243  
 /organism="Ehrlichia chaffeensis"  
 /strain="Jax"  
 /db\_xref="taxon:945"  
 237..1079  
 /codon\_start=1  
 /translation="outer membrane protein p28 precursor"  
 /protein\_id="AAC31546.1"  
 /db\_xref="GI:3414969"  
 /translation="MNYKKVFTSALISLISLPGVSFSDPTGSGINGNFYISGKYM  
 SASFGVFSKEERSTTAGVGLKQDWDGSAISHTPENITVSNYSPKYNPFLGP  
 AGAIGYMDGPRIELEYETFDVKNQNSYKNRAHYCALSHSSSTSSAGDKFV  
 FLKNEGLLDVSFMLNACDVISEGIPSPYICVIGIDLVSMFEATPKLISYQKGL  
 SYSTSPETSVFIGGHHKVIQNEFRDIPAVIPNGSLTGNHPAIVTLDVCHFIGELGG  
 RFAT"

BASE COUNT 396 a 197 c 233 g 417 t

ORIGIN  
 Alignment Scores:  
 Pred. No.: 3,43e-29 Length: 1243  
 Score: 449.50 Matches: 111  
 Percent Similarity: 53.75% Conservative: 54  
 Best Local Similarity: 36.16% Mismatches: 101  
 Query Match: 30.05% Indels: 41  
 DB: 1 Gaps: 12

US-10-062-624-40 (1-293) x AF077733 (1-1243)

Qy 1 MetAsnAsnLysLeuLysPheThrIleAsnThrValLeuValCysLeuLeu---Ser 19  
 Db 237 ATGAATTCACAAAAGTTTTC-----ATAACAAGTGCATTGATATCTATATCTTCT 290  
 Qy 20 LeuProAsnIleSerSerSer-----LysAlaIleAsnAsnAlaLysLys 35  
 Db 291 CTACTGGAGTATCATTTTCTGACCCACACAGTAGTGTATTAACGGCAAT----- 341  
 Qy 36 TyrTyrGlyLeuTyrIleSerGlyClnTyrLysProSerValSerValPheSerAsn 55  
 Db 342 -----TTCTATATCAGTGAATAATATATGCAAGCGCTTCGCATTTTGGGGTGT 392  
 Qy 56 SerValLysGluThrAsnValIleThrLysAsnLeuIleAlaLeuLysAspValAsp 75  
 Db 393 TCTGTAAGGAAGAAAGAAAGTACACAGCTGGAGTATTTGGATTGAACAAGATTGGAT 452  
 Qy 76 SerlleGluThrLysThrAspAlaSerValGlyLysSerAsnProSerAsn---PheThr 94  
 Db 453 GGC-----AGTGCATATCTCACACACCCACCCAGAAAATATATTTCACT 494  
 Qy 95 IlePro---TyrThrAlaValPheClnAspAsnSer---ValAsnPheAsnGlyThrIle 112  
 Db 495 GTTTCAAATATTTCGTTTAAATATGAAATAAACCCATTTTAGGATTTTCAGGAGCTATT 554  
 Qy 113 GlyTyrThrPheAlaGluGlyThrArgValGluIleGluGlySerTyrGluGluPheAsp 132  
 Db 555 GGTACTCAATG---GATGGCCCAAGATAGATGCTTGAAGTATCTTATGAGACATTCGAT 611  
 Qy 133 ValLysAsnProGlyGlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArg 152  
 Db 612 GTAAAAATCAAGGTAAACAGTACAGAACGAGCGCATAGTATTGTGCTTATCTCAT 671  
 Qy 153 GluMetLysGlyAsnSerPheThrProLysGluLysValSerAsnSerllePheHisThr 172  
 Db 672 CACATTCAGGAACAAGCATGTCG-----AGTGCAGGTGATAAATTTGTT 716



QY	173	VaImeArghAnspGlyLeuSerlleSeValleValAsnValCysTyrAspPhe	192
DB	717	TTTTCTAAAAAATGAAGGATTTACTTGGAGTATCGTTTCATGCTGAACGATGCTATGATGA	776
QY	193	SerLeuAsnLeuSerlleSerProTyrIleCysGlyAlaGlyValAspAlaIle	212
DB	777	ATAAGTGAAGGAATACCTTTTCTCTTACATATGTTAGTAGTATCGGACTGATTTAGTA	836
QY	213	GluphePheValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyr	232
DB	837	TCCATGTTTGAAGCTACAAATCTTAAATTTCTTACCAAGGAAGTAGGTTTAAGCTAC	896
QY	233	SerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsn	252
DB	897	TCATTAAGCCCAAGAACTTCTGTGTTTATTGGTGGACATTTTCATTAAGGTGATAGGAAC	956
QY	253	GlnPheLysAsnLeuAsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThr	272
DB	957	GAATTTAGATATATCTT-----GCTGTATACCTAATGATGATCA	995
QY	273	Ser-----AlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGlu	286
DB	996	AGTCTTACAGGAACACCCCTGCAATAGTAACTAGGACGTATGCCACITTTGTTATAGAG	1055
QY	287	IleGlyAlaArgLeuThrPhe	293
DB	1056	CTTGGAGGAAGTTTGCTTTC	1076
RESULT	35		
AF077732			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REMARK			
COMMENT			
FEATURES			
source			
CDS			

```

|||||
995 AGTCTTACAGGAACACCCTGCAATAGTAACACTGGAGCTATGCCACTTTGGTATAGAG 1054
Qy
287 IleGlyAlaArgLeuThrPhe 293
Db
1055 CTGGAGGAGGTTTGCTTTC 1075

RESULT 36
AF3933394
LOCUS
DEFINITION Ehrlichia chaffeensis strain V8 outer membrane protein p28 gene,
complete cds.
ACCESSION AF3933394
VERSION AF3933394.1 GI:15991541
KEYWORDS
SOURCE
ORGANISM Ehrlichia chaffeensis.
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE
1 (bases 1 to 1307)
Long, S.W., Zhang, X.F., Qi, H., Standaert, S., Walker, D.H. and Yu, X.J.
Antigenic variation of Ehrlichia chaffeensis resulting from
differential expression of the 28-kilodalton protein gene family
INFECTION. Immun. 70 (4), 1824-1831 (2002)
JOURNAL
MEDLINE 21893092
PUBMED 11895944
REFERENCE
2 (bases 1 to 1307)
Yu, X.-J. and Zhang, X.-F.
Direct Submission
JOURNAL
SUBMITTED (20-JUN-2001) Pathology, University of Texas Medical
Branch, 301 Univ. Blvd., Galveston, TX 77555-0609, USA
FEATURES
source
1..1307
/organism="Ehrlichia chaffeensis"
/strain="V8"
/db_xref="taxon:945"
CDS
236..1078
/note="p28-19; 28 kDa outer membrane protein"
/codon_start=1
/transl_table=11
/product="outer membrane protein p28"
/protein_id="AAL12924.1"
/db_xref="GI:15991542"
/translation="MNYKKVFITSLISLSPGVSPDPTSGINGNFVISKYMP
SASHGVPSAKEERTAGVFLGKODWGSALISHTPENIFVSNYPKYNPPFLGF
ACATGSDMGPRILEVSEYTFDVKNQNSYKNEAHRYCALSHSSGTSMSAGDKFV
FLKNEGLDVSMNLNACVDVISEGIPFPSPYICVIGITDLVSMFEATNPKISYQKGLG
SYSISPSTVFIGHFHVKVIGNEPRDIPAVIPSGSLGNHPAIVTLVDVCHGIELGG
RPAF"
BASE COUNT 420 a 212 c 245 g 430 t
ORIGIN

Alignment Scores:
Pred. No.: 3,648-29 Length: 1307
Score: 449.50 Matches: 111
Percent Similarity: 53.7% Conservative: 54
Best Local Similarity: 36.1% Mismatches: 101
Query Match: 30.0% Indels: 41
DB: 12 Gaps: 12

US-10-062-624-40 (1-293) x AF3933394 (1-1307)

Qy 1 MetAsnAsnLysLeuLysPheThrIleLeuAsnThrValLeuValCysLeuLeu---Ser 19
Db 236 ATGAATTACAAAAAGTTTTC-----ATAACAAGTGCAATGATATACATAATCTCTCT 289
Qy 20 LeuProAsnIleSerSerSer-----LysAlaIleAsnAsnAlaLysLys 35
Db 290 CTACCTGGATGATCATTTTCTGACCCACACAGTAGTGGTATTAAACGGCAAT----- 340
Qy 36 TyrTyrGlyLeuTyrIleSerGlyGlnTyrLysProSerValSerValPheSerAsnPhe 55
Db 341 -----TTCTATATCATGTCGAAAAATATATATGCAACGCGCTTCGATTTTGGGTGTTT 391

```

```

Qy 56 SerValLysGluThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValasp 75
Db 392 TCTGCTAAGGAAGAAAGAACAGTACACAGCTGGAGTATTGGATTGAAGCAAGATTGGAT 451
Qy 76 SerIleGluThrLysThrAspAlaSerValGlyLysSerAsnProSerAsn---PheThr 94
Db 452 GGC-----AGTGCAATATCTCACACACCCACCCAGAAATATATTACTACT 493
Qy 95 IlePro---TyrThrAlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIle 112
Db 494 GTTTCAAATATTCTGTTTAAATATGAAATAACCACTTTTATAGGATTTGCAGGAGCTATT 553
Qy 113 GlyTyrThrPheAlaGluGlyThrArgValGluIleGluLysSerTyrGluLysPheAsp 132
Db 554 GGTTACTCAATG---GATGCCCAAGAAATAGAGCTTGAAGTATCTTATAGACATTCGAT 610
Qy 133 ValLysAsnProGlyGlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArg 152
Db 611 GTAAAAATCAAGTAAACAGTTACACAGACGACGATAGGTATTGCTTTATCTCAT 670
Qy 153 GluMetLysGlyAsnSerPheThrProLysGluLysValSerAsnSerIlePheHisThr 172
Db 671 CACAGTTTCAGGAACAACAGCATGTCG-----AGTGCAGGTGATAAAATTTGTT 715
Qy 173 ValMetArgAsnAspGlyLeuSerIleSerValIleValAsnValCysTyrAspPhe 192
Db 716 TTTCTAAAAAATGAAGATTACTTTCAGCTATCGTTTCATGCTGAACGCGATGATGATGTA 775
Qy 193 SerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIle 212
Db 776 ATAAGTGAAGGAATACCTTTTCTCCTTACATATGTAGTATCGGAGCTGATTTAGTA 835
Qy 213 GluPhePheAspValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyr 232
Db 836 TCCATGTTTGAAGCTACAAATCTTAAATTTCTTACCAAGGAAAGTAGTTAGTTAAGCTAC 895
Qy 233 SerLeuProSerAsnIleSerLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsn 252
Db 896 TCTATAAGCCCGAGAAACTTCTGTGTTTATTGGTGGACATTTTCATAGGGTATAGGAAC 955
Qy 253 GlnPheLysAsnLeuAsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThr 272
Db 956 GAATTTAGATATATCTCT-----GCTGTAATACCTAGTGGATCA 994
Qy 273 Ser-----AlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGlu 286
Db 995 AGTCTTACAGGAACACCCTGCAATAGTAACACTGGACGATGCCACTTTGGTATAGAG 1054
Qy 287 IleGlyAlaArgLeuThrPhe 293
Db 1055 CTGGAGGAGGTTTGCTTTC 1075

RESULT 37
AF3933390
LOCUS
DEFINITION Ehrlichia chaffeensis strain V4 outer membrane protein p28 gene,
complete cds.
ACCESSION AF3933390
VERSION AF3933390.1 GI:15991533
KEYWORDS
SOURCE
ORGANISM Ehrlichia chaffeensis.
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
REFERENCE
1 (bases 1 to 1309)
Long, S.W., Zhang, X.F., Qi, H., Standaert, S., Walker, D.H. and Yu, X.J.
Antigenic variation of Ehrlichia chaffeensis resulting from
differential expression of the 28-kilodalton protein gene family
INFECTION. Immun. 70 (4), 1824-1831 (2002)
JOURNAL
MEDLINE 21893092
PUBMED 11895944
REFERENCE
2 (bases 1 to 1309)

```

AUTHORS Yu.X.-J., Zhang,X.-F. and Walker,D.H.  
TITLE Direct Submission  
JOURNAL Submitted (20-JUN-2001) Pathology, University of Texas Medical Branch, 301 Univ. Blvd., Galveston, TX 77555-0609, USA  
FEATURES Location/Qualifiers  
source 1. 1309  
/organism="Ehrlichia chaffeensis"  
/strain="V4"  
/db\_xref="taxon:945"  
238. 1080  
/note="p28-19; 28 kDa outer membrane protein"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane protein p28"  
/protein\_id="AAL12920.1"  
/db\_xref="GI:15991534"  
/translation="MNTKKVFITSAISLISLPGVSDPTGSGINGNFVISKYMP  
SASHFVFSKEERSTAGVGLKQDWGSAISHTPENLFTVSNYSFKYENPPFLGF  
AGAIGYMDGPRIELEYSETFDVKNCGSNYKNEAHRYCALSHSSGTSMSAGDKFV  
FLKNEGLDVSFLNACDYVISEGIPSPYICVGIGTDLVSFEATPNPKISYQKGLG  
SYSISPSTSVFIGGHFKVIGNEFRDIPAVIPSGSSLTGNHPAIVTLVDVCHFGIELGG  
RFAP"

BASE COUNT 420 a 212 c 247 g 430 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3.65e-29 Length: 1309  
Score: 449.50 Matches: 111  
Percent Similarity: 53.75% Conservative: 54  
Best Local Similarity: 36.16% Mismatches: 101  
Query Match: 30.05% Indels: 41  
DB: 1 Gaps: 12

US-10-062-624-40 (1-293) x AF393390 (1-1309)

Qy 1 MethAsnLysLeuLysPheThrIleLeaThrValLeuValCysLeuLeu---Ser 19  
Db 238 ATGAATTACAAAAGTTTTC-----ATAACAAGTCGATTGATCACTAATATCTCT 291  
Qy 20 LeuProAsnLysSerSer-----LysAlaIleAsnAsnAlaLysLys 35  
Db 292 CTACTGGGATCATTTCTGACCCACAGTAGTGGTATTACGCCAT----- 342  
Qy 36 TyrTyrGlyLeuTyrLysGlyGlnTyrLysProSerValSerValPheSerAsn 55  
Db 343 -----TTCTATATCAGTGGAAATATATATGCAAGCGTTCGCATTTTGGGGTGT 393  
Qy 56 SerValLysGluThrAsnValIleThrLysAsnLeuLeaLysLysAspValAsp 75  
Db 394 TCTGCTAAGGAAGAAAGAGTACACAGCTGGAGTATTGGATTGAAGCAAGATTGGAT 453  
Qy 76 SerIleGluThrLysThrAspAlaSerValGlyTyrSerAsnProSerAsn---PheThr 94  
Db 454 GGC-----AGTGCAATATCTCACACCACCCAGCAATATATCTCACT 495  
Qy 95 IlePro---TyrThrAlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThrIle 112  
Db 496 GTTTCAAAATATTCGTTTAAATATGAAATAACCACTTTTATGAGATTTCGAGGAGCTATT 555  
Qy 113 GlyTyrThrPheAlaGluGlyThrArgValGluLeuGlySerTyrGluGluPheAsp 132  
Db 556 GGTACTCAATG---GATGCCCAAGAAATAGAGTTTGAAGTATCTTATGACATTCGAT 612  
Qy 133 ValLysAsnProGlyGlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAlaArg 152  
Db 613 GTAAAAATCAAGTACAGTACAGTACAGCAAGCAAGCAAGTATGTTGCTTTATCTCAT 672  
Qy 153 GluMetLysGlyAsnSerPheThrProLysGluLysValSerAsnSerIlePheHisThr 172  
Db 673 CACAGTTTCAGGAACAAGCATGCG-----AGTGCAGGTGATAAATTTGTT 717  
Qy 173 ValMetArgAsnAspGlyLeuSerIleLeaSerValIleValAsnValCysTyrAspPhe 192

Db 718 TTTCTAAAAAATGAAGGATTACTTGTACGATCGTTTCATGCTGAACGATCGTATGATGTA 777  
Qy 193 SerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGlyValAspAlaIle 212  
Db 778 ATAGTGAAGGAATACCTTTTCTCTTACATATGTGTAGTATCGGACTGATTAGTA 837  
Qy 213 GluPhePheAspValLeuHisIleLysPheAlaTyrGlnSerLysLeuGlyIleAlaTyr 232  
Db 838 TCCATGTTTGAAGCTACAAATCTCTAAATTTCTTACCAAGAAAGTTAGGTTTAAAGTAC 897  
Qy 233 SerLeuProSerAsnLysLeuPheAlaSerLeuTyrTyrHisLysValMetGlyAsn 252  
Db 898 TCTATAAGCCAGAACTTCTGTTTATGTTGGTGGACATTTTATAGGTGATAGGAAC 957  
Qy 253 GlnPheLysAsnLeuAsnValGlnHisValAlaGluLeuAlaSerIleProLysIleThr 272  
Db 958 GAATTTAGATATTCCT-----GCTGTAAATACCTAGTGGATCA 996  
Qy 273 Ser-----AlaValAlaThrLeuAsnIleGlyTyrPheGlyGlyGlu 286  
Db 997 AGTCTTACAGAAACCACTGCAATAGTAACACTGGACGTATGCCACTTTGGTATAGAG 1056  
Qy 287 IleGlyAlaArgLeuThrPhe 293  
Db 1057 CTGGAGGAAGTTTGTCTTTC 1077  
RESULT 38  
AF368008 825 bp DNA linear BCT 06-NOV-2001  
LOCUS Cowdria ruminantium isolate Mara87/7 major antigenic protein MAP1  
DEFINITION gene, partial cds.  
ACCESSION AF368008  
VERSION AF368008.1 GI:15429327  
KEYWORDS  
SOURCE Ehrlichia ruminantium.  
ORGANISM Ehrlichia ruminantium.  
REFERENCE 1 (bases 1 to 825)  
AUTHORS Allsopp,M.T., Dorfling,C.M., Maillard,J.C., Bensaid,A., Haydon,D.T., van Heerden,H. and Allsopp,B.A.  
TITLE Ehrlichia ruminantium major antigenic protein gene (map1) variants are not geographically constrained and show no evidence of having evolved under positive selection pressure  
J. Clin. Microbiol. 39 (11), 4200-4203 (2001)  
JOURNAL MEDLINE 21539003  
PUBMED 11682561  
REFERENCE 2 (bases 1 to 825)  
AUTHORS Allsopp,M.T., Dorfling,C., Maillard,J.-C., Bensaid,A., Van Heerden,H. and Allsopp,B.A.  
TITLE Direct Submission  
JOURNAL Submitted (05-APR-2001) Molecular Biology, Onderstepoort Veterinary Institute, Old Soutpansberg Road, Onderstepoort, Gauteng 0110, South Africa  
FEATURES Location/Qualifiers  
source 1. 825  
/organism="Ehrlichia ruminantium"  
/isolate="Mara87/7"  
/db\_xref="taxon:779"  
CDS <1..825  
/codon\_start=1  
/transl\_table=11  
/product="major antigenic protein MAP1"  
/protein\_id="AAK98148.1"  
/db\_xref="GI:15429328"  
/translation="FITSLISLVSLPGVSDPTQEDSNPAGSVISAKYMPATASH  
FGWISKEDSKNTQTVFGKKQDWGVTSPDSNTNSTIFTEDKDYFRYNNPFLGFA  
GATGYSMGRIEFYFETFDVKNFGNKKNNAHMYCALDTAAQNSTNGAGLTTSYM  
VKNEINTNISLMLNACDYVISEGIPSPYICVGIGTDLVSNATPNPKLSYQKGLGS  
YSINSEASIFLGHFHRVIGNEFKDITATLKIFTKSGISNPGFASATLDVCHFGIEI"  
BASE COUNT 293 a 139 c 138 g 255 t  
ORIGIN

RESULT 39  
AF368010  
LOCUS

```

Qy 89 AsnProSerAsnPhetThr---IleProTyrThrAlaValPheGlnAspAsnSer---Val 106
Db 244 -----TCACCTTTTACTGAAAAAGATTATCTTTTCAGATATGAAAAACAATCCGTTTTTA 297
Qy 107 AsnPhetAsnGlyThrIleGlyTyrThrPheAlaGluGlyThrArgValGluIleGluGly 126
Db 298 GGTTTTCTGCTGAGCAATGGGTACTCAATG---AATGACCAACAAGATAGATTCGAAGTA 354
Qy 127 SerTyrGluGluPheAspValIleAsnProGlyGlyTyrThrLeuSerAspAlaTyrArg 146
Db 355 TCCTATGAACCTTTTGATGTAATAAAACCCAGGTGCGCACTACAAATAATGATGCACATG 414
Qy 147 TyrPheAlaLeu-----AlaArgGluMetLysGlyAsnSerPheThrProLysGluLys 164
Db 415 TACTGTGCTTTAGATACAGACACAGAGTCTACTAATGTTGTCACAACA-----462
Qy 165 ValSerAsnSerIlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSerVal 184
Db 463 TTAGCTTCATCTGTT-----GTGATAAAAAATGAAATTTAACAATATATATCACTA 513
Qy 185 IleValAsnValCysTyrAspPheSerLeuAsnAsnLeuSerIleSerProTyrIleCys 204
Db 514 ATGTTAATCGCTGTATGATATATGCTTGTGATGGAATGCCAGTTTCTCCATATGTATGT 573
Qy 205 GlyGlyAlaGlyValAspAlaIleGluPhePheAspValLeuHisIleLysPheAlaTyr 224
Db 574 GCAGGTATGCTGCTGATTTAGTCTAGTAATTAATGCTACAAATCTCAAAATATATCTTAT 633
Qy 225 GlnSerLysLeuGlyIleAlaTyrSerLeuProSerAsnIleSerLeuPheAlaSerLeu 244
Db 634 CAAGAAAGCTAGGATCAATGTTACTCAATCTCGAGGCTTCTATCTTTATCGGTGGA 693
Qy 245 TyrTyrHisLysValMetGlyAsnGlnPheLysAsnLeuAsnValGlnHisVal-----262
Db 694 CATTTCCATAGAGTATAGTAATGAAATTTAAGATATGCTACTTTTAAATAATATTACC 753
Qy 263 -----AlaGluLeuAlaSerIleProLysIleThrSerAlaValAlaThrLeuAsnIle 280
Db 754 GCAACTAATAAGTATCTACTGTAGCCCAATCCAGGTTTGTGCATCAGCAACACTTGATGTT 813
Qy 281 GlyTyrPheGlyGlyIle 287
Db 814 TGTCACTTTGTTATAGAAATT 834

RESULT 40
AF393393 LOCUS
DEFINITION Ehrlichia chaffeensis strain V7 outer membrane protein p28 gene,
complete cds.
ACCESSION AF393393
VERSION AF393393.1 GI:15991539
KEYWORDS
SOURCE ORGANISM
Ehrlichia chaffeensis.
Ehrlichia chaffeensis
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Anaplasmataceae; Ehrlichia.
1 (bases 1 to 1286)
Long, S.W., Zhang, X.F., Qi, H., Standaert, S., Walker, D.H. and Yu, X.J.
Antigenic variation of Ehrlichia chaffeensis resulting from
differential expression of the 28-kilodalton protein gene family
Infect. Immun. 70 (4), 1824-1831 (2002)
JOURNAL MEDLINE
PUBMED 21893092
REFERENCE 2 (bases 1 to 1286)
AUTHORS Yu, X.-J. and Zhang, X.-F.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2001) Pathology, University of Texas Medical
Branch, 301 Univ. Blvd., Galveston, TX 77555-0609, USA
FEATURES
source
Location/Qualifiers
1..1286
/organism="Ehrlichia chaffeensis"
/strain="V7"

```

```

/db_xref="taxon:945"
243..1073
/notes="p28-19; 28 kDa outer membrane protein"
/codon_start=1
/transl_table=11
/product="outer membrane protein p28"
/protein_id="AAU12923.1"
/db_xref="GI:15991540"
/translation="MNYKVFITSAIISLISLPGVSFSDPTSGINGNFYISGKYP
SAGHFGVFSABERNNTAGVFLQKQWDGSAIPHHTSNTPTFTVSNYFKYENNPFLG
AGAIGYSDMPRIELEVSYETFDQNNYKNEAHRYCALSRATTAADNTVFLKN
EGLDLSFMLNACVDVTSEGIPESPYICAGIGTDLVSMFEVTPKISKQGLGSYI
SPETSFVGVGHFHKVGNRPDRDPAIIPSSSLTGNHPAIVTLDVCHFGIELGRFAP
"
BASE COUNT 428 a 209 c 221 g 428 t
ORIGIN
Alignment Scores:
Pred. No.: 5,27e-29 Length: 1286
Score: 447.50 Matches: 112
Percent Similarity: 52.73% Conservative: 52
Best Local Similarity: 36.01% Mismatches: 94
Query Match: 29.91% Indels: 53
Gaps: 12
US-10-062-624-40 (1-293) x AF393393 (1-1286)
Qy 1 MetAsnAsnLysLeuLysPheThrIleAsnThrValLeuValCysLeuLeu---Ser 19
Db 243 ATGNAATTCAGAAAAGTTTC-----ATACAAGTCGATGATATCACTAATATCTCT 296
Qy 20 LeuProAsnIleSerSerSer-----LysAlaIleAsnAsnAlaLysLys 35
Db 297 CTACCTGGAGTATCATTTCCGACCCCAACAGGTAGTGTATTAAACGCAAT-----347
Qy 36 TyrTyrGlyLeuTyrIleSerGlyGlnTyrLysProSerValSerValPheSerAsnPhe 55
Db 348 -----TTCATATATCAGTGGAAAAATATATGCAAGCGCTTCGCATTTTGGGGTGT 398
Qy 56 SerValLysGluThrAsnValIleThrLysAsnLeuIleAlaLeuLysLysAspValAsp 75
Db 399 TCTGCTAAGGAAGAAAGAAATACACAGCTGGAGTATTGGATTGAAGCAAGATTGGAT 458
Qy 76 SerIle-----GluThrLysThrAspAlaSerValGlyIleSerAsnProSerAsn 92
Db 459 GGCAGTCAATACCTCACATCATAGTAATATCTCAATTTACTGTTTCAAAAC-----509
Qy 93 PheThrIleProTyrThrAlaValPheGlnAspAsnSer---ValAsnPheAsnGlyThr 111
Db 510 -----TACTCATTTAAATATGAAATAATCCATTTCTAGGTTTTCAGGAGCT 557
Qy 112 IleGlyTyrThrPheAlaGluGlyThrArgValGluIleGluGlySerTyrGluGluPhe 131
Db 558 ATTGGCTACTCAATG---GATGGTCCAAGATAGAGTTGAAATGATCTTATGAGACATTT 614
Qy 132 AspValLysAsnProGlyGlyTyrThrLeuSerAspAlaTyrArgTyrPheAlaLeuAla 151
Db 615 GATGTTAAATAATCAAGTAACTAATAAGACGAACTCATAGATATCTAGCTTATATCC 674
Qy 152 ArgGlu-----MetLysGlyAsnSerPheThrProLysGluLysValSerAsnSer 168
Db 675 CGTCAGGCCCAATAGCAGCAGATAATACATTT-----707
Qy 169 IlePheHisThrValMetArgAsnAspGlyLeuSerIleIleSerValIleValAsnVal 188
Db 708 -----GTTTCTTAAATAATGAAAGGCTACTTGACATATCATTTATGTAACGCA 758
Qy 189 CysTyrAspPheSerLeuAsnAsnLeuSerIleSerProTyrIleCysGlyGlyAlaGly 208
Db 759 TGTATGATGTAATAAGTGAAGGATATACCTTTTCTCTTACATATGTGCAAGTATGGT 818
Qy 209 ValAspAlaIleGluPhePheAspValLeuHisIleLysPheAlaTyrGlnSerLysLeu 228

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 10:07:40 ; Search time 72 Seconds  
(without alignments)  
542.256 Million cell updates/sec

Title: US-10-062-624-40  
Perfect score: 1496  
Sequence: 1 MNKKLFTIINTVLCLSL.....AVATLNGYFGGIGARLTF 293

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1496	100.0	293	20	E. canis P30-5 pro
2	1496	100.0	293	23	Ehrlichia canis p2
3	1033	69.1	291	23	Ehrlichia chaffeensis
4	1032.5	69.0	300	23	Ehrlichia chaffeensis
5	1026	68.6	293	23	Ehrlichia chaffeensis
6	1006.5	67.3	298	23	Ehrlichia chaffeensis
7	999.5	66.8	292	20	E. canis P30-7 pro
8	995	66.5	280	20	E. chaffeensis OMP-
9	930	62.2	271	20	E. chaffeensis OMP-
10	923	61.7	299	20	E. canis P30-8 pro

11	789	52.7	285	20	AAV06957	E. chaffeensis OMP-
12	789	52.7	285	23	AAU73408	Ehrlichia chaffeensis
13	786.5	52.6	279	20	AAV06954	E. chaffeensis OMP-
14	779.5	52.1	279	23	AAU73405	Ehrlichia chaffeensis
15	768	51.3	283	20	AAV06955	E. chaffeensis OMP-
16	768	51.3	283	23	AAU73406	Ehrlichia chaffeensis
17	737.5	49.3	275	20	AAV06956	E. chaffeensis OMP-
18	737.5	49.3	275	23	AAU73407	Ehrlichia chaffeensis
19	720	48.1	289	20	AAV06966	E. canis P30-6 pro
20	671	44.9	295	20	AAV06953	E. chaffeensis OMP-
21	671	44.9	295	23	AAU73404	Ehrlichia chaffeensis
22	574	38.4	164	20	AAV06951	E. chaffeensis OMP-
23	519	34.7	272	23	AAU73403	Ehrlichia chaffeensis
24	500.5	33.5	161	20	AAV06969	E. canis P30-9 pro
25	498	33.3	283	20	AAV06944	E. chaffeensis OMP-
26	498	33.3	283	23	AAU96106	Ehrlichia chaffeensis
27	498	33.3	283	23	AAU73413	Ehrlichia chaffeensis
28	490.5	32.8	280	20	AAU96116	E. chaffeensis OMP-
29	470.5	31.5	280	20	AAV06948	Ehrlichia canis p2
30	470.5	31.5	280	23	AAU96110	Ehrlichia chaffeensis
31	470.5	31.5	288	20	AAV06959	E. canis P30 prote
32	468.5	31.3	280	19	AAW51094	Ehrlichia chaffeensis
33	468.5	31.3	280	21	AAV36188	Ehrlichia chaffeensis
34	468.5	31.3	280	22	AAU04198	Variable surface a
35	468.5	31.3	280	23	AAU73417	Ehrlichia chaffeensis
36	464	31.0	271	23	AAU73419	Ehrlichia chaffeensis
37	459	30.7	287	19	AAW51088	Cowdria ruminantii
38	459	30.7	287	21	AAV36182	Cowdria ruminantii
39	459	30.7	287	22	AAU04192	Major antigenic pr
40	456.5	30.5	284	23	AAU96111	Cowdria ruminantii
41	450.5	30.1	280	20	AAV06945	E. chaffeensis OMP-
42	450.5	30.1	280	23	AAU96107	Ehrlichia chaffeensis
43	450.5	30.1	280	23	AAU73414	Ehrlichia chaffeensis
44	448	29.9	284	23	AAU73400	Ehrlichia chaffeensis
45	446.5	29.8	280	21	AAV71479	Ehrlichia canis im

ALIGNMENTS

RESULT 1

AAV06965  
ID AAY05965 standard; Protein; 293 AA.

XX AAY06965;

XX AC

XX 05-JUL-1999 (first entry)

DT XX

DE E. canis P30-5 protein.

DE XX

KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;

KW detection; dog.

XX XX

OS Ehrlichia canis.

XX XX

PN WO9913720-A1.

XX XX

PD 25-MAR-1999

XX XX

XX 18-SEP-1998; 98WO-US19600.

XX XX

PR 19-SEP-1997; 97US-0059353.

XX XX

PA (OHIS ) UNIV OHIO STATE.

XX XX

PI Ohashi N, Rikihisa Y;

XX XX

DR WPI; 1999-254290/21.

DR N-PSDB; AAX34765.

PT Novel Outer-membrane proteins from Ehrlichia chaffeensis and

PT Ehrlichia canis

XX XX

PS Disclosure; Fig 25B; 55pp; English.

XX The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 CC in AAY06943-958. The E. canis proteins form part of the p30 family and  
 CC consist of proteins shown in AAY06959-970. The proteins and genes are  
 CC used to detect E. chaffeensis in patients and E. canis in dogs.

SQ Sequence 293 AA;  
 Query Match 100.0%; Score 1496; DB 20; Length 293;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-135;  
 Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MNKLKFTIINTVLVCLSLPNISSKAINNNAKYYGLYISGQKPSVSVFNSFKET 60  
 DB 1 MNKLKFTIINTVLVCLSLPNISSKAINNNAKYYGLYISGQKPSVSVFNSFKET 60  
 QY 61 NVITKNLALKKDVDSIETKTDSVGSINPSNFTIPYTAVFQDQNSVNFNGTIGYTFABGT 120  
 DB 61 NVITKNLALKKDVDSIETKTDSVGSINPSNFTIPYTAVFQDQNSVNFNGTIGYTFABGT 120  
 QY 121 RVEIGSYEEEDVKNPGGYTILSDAYRYFALAREMKGNSFTPKKVSNSIFHTVNRNDGLS 180  
 DB 121 RVEIGSYEEEDVKNPGGYTILSDAYRYFALAREMKGNSFTPKKVSNSIFHTVNRNDGLS 180  
 QY 181 IISVIVNVCYDFSLNLSISPIYCGGAGVDAIEFFDVLHIKPAYQSKLGIAYSLPSNISL 240  
 DB 181 IISVIVNVCYDFSLNLSISPIYCGGAGVDAIEFFDVLHIKPAYQSKLGIAYSLPSNISL 240  
 QY 241 PASLYYHKMGQNFKNLVQHVLAELASIPKITSATATNLNIGYFGGEIGARLTF 293  
 DB 241 PASLYYHKMGQNFKNLVQHVLAELASIPKITSATATNLNIGYFGGEIGARLTF 293

# RESULT 2

AAU96115  
 ID AAU96115 standard; Protein; 293 AA.  
 AC AAU96115;  
 XX  
 DT 02-JUL-2002 (first entry)  
 XX  
 DE Ehrlichia canis p28-1.  
 XX  
 KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;  
 KW antibacterial.  
 XX  
 OS Ehrlichia canis.  
 XX  
 PN WO200222782-A2.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PF 12-SEP-2001; 2001WO-US28759.  
 XX  
 PR 12-SEP-2000; 2000US-0660587.  
 XX  
 PA (RERE-) RES DEV FOUND.  
 XX  
 PI Walker DH, Yu X, McBride JW;  
 XX  
 XX WPI; 2002-351882/38.  
 DR N-PSDB; ABK68875.  
 XX  
 PT New recombinant homologous 28 kilodalton immunodominant protein from  
 PT Ehrlichia canis, useful for treating Ehrlichia canis infections -  
 XX  
 PS Claim 16; Figure 13; 105pp; English.  
 XX  
 CC The invention relates to a recombinant homologous 28 kDa immunodominant  
 CC protein, p28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably

CC dispersed in a pharmaceutically acceptable carrier, is useful for  
 CC inhibiting E. canis infection in a subject. (I) is useful in the  
 CC development of vaccines and serodiagnostics that are particularly  
 CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118  
 CC represent the 28-kDa antigen amino acid sequences of the invention.

SQ Sequence 293 AA;

Query Match 100.0%; Score 1496; DB 23; Length 293;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-135;  
 Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MNKLKFTIINTVLVCLSLPNISSKAINNNAKYYGLYISGQKPSVSVFNSFKET 60  
 DB 1 MNKLKFTIINTVLVCLSLPNISSKAINNNAKYYGLYISGQKPSVSVFNSFKET 60  
 QY 61 NVITKNLALKKDVDSIETKTDSVGSINPSNFTIPYTAVFQDQNSVNFNGTIGYTFABGT 120  
 DB 61 NVITKNLALKKDVDSIETKTDSVGSINPSNFTIPYTAVFQDQNSVNFNGTIGYTFABGT 120  
 QY 121 RVEIGSYEEEDVKNPGGYTILSDAYRYFALAREMKGNSFTPKKVSNSIFHTVNRNDGLS 180  
 DB 121 RVEIGSYEEEDVKNPGGYTILSDAYRYFALAREMKGNSFTPKKVSNSIFHTVNRNDGLS 180  
 QY 181 IISVIVNVCYDFSLNLSISPIYCGGAGVDAIEFFDVLHIKPAYQSKLGIAYSLPSNISL 240  
 DB 181 IISVIVNVCYDFSLNLSISPIYCGGAGVDAIEFFDVLHIKPAYQSKLGIAYSLPSNISL 240  
 QY 241 PASLYYHKMGQNFKNLVQHVLAELASIPKITSATATNLNIGYFGGEIGARLTF 293  
 DB 241 PASLYYHKMGQNFKNLVQHVLAELASIPKITSATATNLNIGYFGGEIGARLTF 293

# RESULT 3

AAU73409  
 ID AAU73409 standard; Protein; 291 AA.  
 AC AAU73409;  
 XX  
 DT 12-MAR-2002 (first entry)  
 XX  
 DE Ehrlichia chaffeensis outer membrane protein p28-10.  
 XX  
 KW Ehrlichia; outer membrane protein; p28; antibiotic; vaccine.  
 XX  
 OS Ehrlichia chaffeensis.  
 XX  
 PN WO200183699-A2.  
 XX  
 PD 08-NOV-2001.  
 XX  
 PF 01-MAY-2001; 2001WO-US13997.  
 XX  
 PR 01-MAY-2000; 2000US-201035P.  
 XX  
 PA (RERE-) RES DEV FOUND.  
 XX  
 PI Walker DH, Yu X;  
 XX  
 XX WPI; 2002-066527/09.  
 DR  
 XX  
 PT Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated  
 PT p28 useful as a vaccine against Ehrlichia chaffeensis -  
 XX  
 PS Claim 10; Figure 2; 97pp; English.  
 XX  
 CC The invention relates to isolated and purified 28-kDa outer membrane  
 CC proteins (p28-1 to p28-21) of Ehrlichia chaffeensis. p28 proteins  
 CC are encoded by a 28kDa outer membrane protein multigene family. p28  
 CC proteins are useful as a vaccine against E. chaffeensis. DNA encoding p28  
 CC is useful for transfecting a host cell. AAU73400-AAU73420 represent  
 CC Ehrlichia chaffeensis p28 outer membrane proteins of the invention.



SQ Sequence 291 AA;  
 Query Match 69.1%; Score 1033; DB 23; Length 291;  
 Best Local Similarity 68.6%; Pred. No. 7.3e-91;  
 Matches 201; Conservative 34; Mismatches 56; Indels 2; Gaps 2;  
 QY 1 MNNKLFTTINTVLCCLSLPNISSKAINNNAKKYGLYISGQYKPSVSPFSVKET 60  
 DB 1 MKKKNFITITILVLCCLSLNASLNTTNSSTKKQFGLYVSGQYKPSVSPFSVKET 60  
 QY 61 NVITKNIALKKDQVDSIETKTASVGSINPSNFTPIYTAVFODNSWNGTIGYTFASGT 120  
 DB 61 NFTPKYLAALKKDINSVEFDDSVTAGISYPLNFTPIYTAVFODNISNFGAIGYTFVEGP 120  
 QY 121 RVEIGSYBEFDVKNPGGYT-LSDAYRYFALAREMKGNSFTPKKVSN-----SIFHTV 173  
 DB 121 RVEIGSYBEFDVKNPGGYTQDAYRYFALARDIDSPTSPKNTSHDGNSSYKVYHTV 180  
 QY 174 MRNDGLSIISVINVCYDFSLNNLSISPYICGAGVDAIEFFDVLHIKPAYOSKLGIAYS 233  
 DB 181 MKNEGSIISIMVNGCYDFSSDNLSTLPYVCGGIGVNAIEFFDVLHIKPAYOSKLGITVP 240  
 QY 234 LPSNISLASYHVKMGNOFKNLNVQHVLAELASIPKITSAVATNLNIGYFGGEIGARLTF 293  
 DB 241 LSSNLVFAAGYHQVMGNQFKNLNVQHVLAELNDAPKVTSAVATLIDIGYFGGEIGARLTF 300  
 RESULT 5  
 AAU73412  
 ID AAU73412 standard; Protein; 293 AA.  
 XX AC AAU73412;  
 XX AC  
 XX 12-MAR-2002 (first entry)  
 DT DT  
 DE Ehrlichia chaffeensis outer membrane protein P28-13.  
 XX KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.  
 XX OS Ehrlichia chaffeensis.  
 XX XX  
 PN WO200183699-A2.  
 XX PD 08-NOV-2001.  
 XX XX  
 XX 01-MAY-2001; 2001WO-US13997.  
 XX PF  
 XX 01-MAY-2000; 2000US-201035P.  
 PR  
 XX (RERE-) RES DEV FOUND.  
 XX PA Walker DH, Yu X;  
 XX PI  
 XX WPI; 2002-066527/09.  
 DR  
 XX Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated P28 useful as a vaccine against Ehrlichia chaffeensis -  
 XX PS Claim 10; Figure 2; 97pp; English.  
 XX CC The invention relates to isolated and purified 28-kDa outer membrane proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins are encoded by a 28kDa outer membrane protein multigene family. P28 proteins are useful as a vaccine against E. chaffeensis. DNA encoding P28 is useful for transfecting a host cell. AAU73400-AAU73420 represent Ehrlichia chaffeensis P28 outer membrane proteins of the invention.  
 XX SQ Sequence 293 AA;  
 Query Match 68.6%; Score 1026; DB 23; Length 293;  
 Best Local Similarity 66.0%; Pred. No. 3.5e-90;  
 Matches 194; Conservative 40; Mismatches 58; Indels 2; Gaps 2;  
 QY 1 MNNKLFTTINTVLCCLSLPNISSKAINNNAKKYGLYISGQYKPSVSPFSVKET 60  
 DB 1 MNKNKFFITITAMVLCCLLPGLISFSETINNSAKQPGLYISGQYKPSVSPFSVKET 60  
 QY 61 NVITKNIALKKDQVDSIETKTASVGSINPSNFTPIYTAVFODNSWNGTIGYTFASGT 120  
 DB 61 NFTPKYLAALKKDINSVEFDDSVTAGISYPLNFTPIYTAVFODNISNFGAIGYTFVEGP 120  
 QY 121 RVEIGSYBEFDVKNPGGYT-LSDAYRYFALAREMKGNSFTPKKVSN-----SIFHTV 173  
 DB 121 RVEIGSYBEFDVKNPGGYTQDAYRYFALARDIDSPTSPKNTSHDGNSSYKVYHTV 180  
 QY 174 MRNDGLSIISVINVCYDFSLNNLSISPYICGAGVDAIEFFDVLHIKPAYOSKLGIAYS 233  
 DB 181 MKNEGSIISIMVNGCYDFSSDNLSTLPYVCGGIGVNAIEFFDVLHIKPAYOSKLGITVP 240  
 QY 234 LPSNISLASYHVKMGNOFKNLNVQHVLAELASIPKITSAVATNLNIGYFGGEIGARLTF 293  
 DB 241 LSSNLVFAAGYHQVMGNQFKNLNVQHVLAELNDAPKVTSAVATLIDIGYFGGEIGARLTF 300  
 RESULT 4  
 AAU73411  
 ID AAU73411 standard; Protein; 300 AA.  
 XX AC AAU73411;  
 XX AC  
 XX 12-MAR-2002 (first entry)  
 DT DT  
 DE Ehrlichia chaffeensis outer membrane protein P28-12.  
 XX KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.  
 XX OS Ehrlichia chaffeensis.  
 XX XX  
 PN WO200183699-A2.  
 XX PD 08-NOV-2001.  
 XX XX  
 XX 01-MAY-2001; 2001WO-US13997.  
 XX PF  
 XX 01-MAY-2000; 2000US-201035P.  
 PR  
 XX (RERE-) RES DEV FOUND.  
 XX PA Walker DH, Yu X;  
 XX PI  
 XX WPI; 2002-066527/09.  
 DR  
 XX Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated P28 useful as a vaccine against Ehrlichia chaffeensis -  
 XX PS Claim 10; Figure 2; 97pp; English.  
 XX CC The invention relates to isolated and purified 28-kDa outer membrane proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins are encoded by a 28kDa outer membrane protein multigene family. P28 proteins are useful as a vaccine against E. chaffeensis. DNA encoding P28 is useful for transfecting a host cell. AAU73400-AAU73420 represent Ehrlichia chaffeensis P28 outer membrane proteins of the invention.  
 XX SQ Sequence 300 AA;  
 Query Match 69.0%; Score 1032.5; DB 23; Length 300;  
 Best Local Similarity 66.7%; Pred. No. 8.6e-91;  
 Matches 200; Conservative 35; Mismatches 58; Indels 7; Gaps 2;  
 QY 1 MNNKLFTTINTVLCCLSLPNISSKAINNNAKKYGLYISGQYKPSVSPFSVKET 60  
 DB 1 MNKNKFFITITAMVLCCLLPGLISFSETINNSAKQPGLYISGQYKPSVSPFSVKET 60  
 QY 61 NVITKNIALKKDQVDSIETKTASVGSINPSNFTPIYTAVFODNSWNGTIGYTFASGT 120  
 DB 61 NFTPKYLAALKKDINSVEFDDSVTAGISYPLNFTPIYTAVFODNISNFGAIGYTFVEGP 120  
 QY 121 RVEIGSYBEFDVKNPGGYT-LSDAYRYFALAREMKGNSFTPKKVSN-----SIFHTV 173  
 DB 121 RVEIGSYBEFDVKNPGGYTQDAYRYFALARDIDSPTSPKNTSHDGNSSYKVYHTV 180  
 QY 174 MRNDGLSIISVINVCYDFSLNNLSISPYICGAGVDAIEFFDVLHIKPAYOSKLGIAYS 233  
 DB 181 MKNEGSIISIMVNGCYDFSSDNLSTLPYVCGGIGVNAIEFFDVLHIKPAYOSKLGITVP 240  
 QY 234 LPSNISLASYHVKMGNOFKNLNVQHVLAELASIPKITSAVATNLNIGYFGGEIGARLTF 293  
 DB 241 LSSNLVFAAGYHQVMGNQFKNLNVQHVLAELNDAPKVTSAVATLIDIGYFGGEIGARLTF 300

Db 61 NVPTKQLIALKKDINSVAVGSAATTGNSPGNFTTPTAERQDNVANFENGAVGYSPDLSL 120  
Qy 121 RVEIGSYEEFDVKNPQGYT-LSDAYRYPALAREMKGNSFTPKKEKVSNSIFHTVMRNDGL 179  
Db 121 RIEIEGPEHEFDVKNPQGYTQKDAYRYFALARDLKDGFEPKAE-DTGVTHTVMKNDGL 179  
Qy 180 SIISVIVNVVCDPSLNNLSISPIYICGGAGVDAIEFFDLVHLIKFAYOSKLGIAYSIPSNIS 239  
Db 180 SILSTMVNVVCDPSLNNLSISPIYICGGAGVDAIEFFDLVHLIKFAYOSKLGIAYSIPSNIS 239  
Qy 240 LFPASLYHHKVMGNQFNLMVHVAELASIPKITSATATINIGYFGGEGICARLTF 293  
Db 240 LFLDGYHHQVIGNQFNLMVHVAELASIPKITSATATINIGYFGGEGIRFTF 293  
RESULT 6  
ID AAU73410 standard; Protein; 298 AA.  
XX AAU73410;  
AC AAU73410;  
XX 12-MAR-2002 (first entry)  
XX Ehrlichia chaffeensis outer membrane protein P28-11.  
XX Ehrlichia chaffeensis outer membrane protein; P28; antibiotic; vaccine.  
XX Ehrlichia chaffeensis.  
XX WO200183699-A2.  
XX 08-NOV-2001.  
XX 01-MAY-2001; 2001WO-US13997.  
XX 01-MAY-2000; 2000US-201035P.  
XX (RERE-) RES DEV FOUND.  
XX Walker DH, Yu X;  
XX WPI; 2002-066527/09.  
XX Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated  
PT P28 useful as a vaccine against Ehrlichia chaffeensis -  
XX Claim 10; Figure 2; 97pp; English.  
XX The invention relates to isolated and purified 28-kDa outer membrane  
CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins  
CC are encoded by a 28kDa outer membrane protein multigene family. P28  
CC proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28  
CC is useful for transfecting a host cell. AAU73400-AAU73420 represent  
CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.  
XX Sequence 298 AA;  
SQ Query Match 67.3%; Score 1006.5; DB 23; Length 298;  
Best Local Similarity 65.2%; Pred. No. 2.7e-88;  
Matches 193; Conservative 40; Mismatches 56; Indels 7; Gaps 3;  
Qy 1 MNKCLKFTIINTVLVCLSLPNISSKAINNNAKYYGLYISGQYKPSVSFNSFSVKET 60  
Db 1 MNHSMLEFTIGALISLSLNPVSFGIINNANN-LGIYISGQYKPSVSFNSFSVKET 59  
Qy 61 NVITKNLIALKKDVSIEKTDASVGISNPSNFTPTATVAFQDNSVNFNGTIGYTFACET 120  
Db 60 NFFTQQLVALKKDIDSDISTNADSGINNPQNTPTPTPKFQDNRASFGALGFFYARGL 119  
Qy 121 RVEIGSYEEFDVKNPQGYT-LSDAYRYPALAREMKGNSFTPKKEVSNSIFHTVM 174  
Db 120 RLEMEGSYEEFDVKNPQGYTKVDAYRYPALAREMQSGTQCPKHETSGIQPHGIYHTVM 179

Qy 175 RNDGLSIISVIVNVVCDPSLNNLSISPIYICGGAGVDAIEFFDLVHLIKFAYOSKLGIAYSL 234  
Db 180 RNDGVSISVINGCYNTILSNLPISPYCMVGMDAIQFDSHLIKFAHQSKLGIYFPL 239  
Qy 235 PSNITSLFASLYHHKVMGNQFNLMVHVAELASIPKITSATATINIGYFGGEGICAR 290  
Db 240 SSNVHLPADSYHHKVMGNQFNLMVHVAELASIPKITSATATINIGYFGGEGICAR 295  
RESULT 7  
ID AAU06967 standard; Protein; 292 AA.  
XX AAU06967;  
XX 05-JUL-1999 (first entry)  
XX E. canis P30-7 protein.  
XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
KW detection; dog.  
XX Ehrlichia canis.  
XX WO9913720-A1.  
XX 25-MAR-1999.  
XX 18-SEP-1998; 98WO-US19600.  
XX 19-SEP-1997; 97US-0059353.  
XX (OHIS) UNIV OHIO STATE.  
XX Chashi N, Rikihisa Y;  
XX WPI; 1999-254290/21.  
XX N-PSDB; AAX34767.  
XX Novel outer membrane proteins from Ehrlichia chaffeensis and  
PT Ehrlichia canis  
XX Disclosure; Fig 27B; 55pp; English.  
XX The invention provides isolated outer membrane proteins (OMP) from  
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
CC in AAU06943-958. The E. canis proteins form part of the P30 family and  
CC consist of proteins shown in AAU06959-970. The proteins and genes are  
CC used to detect E. chaffeensis in patients and E. canis in dogs.  
XX Sequence 292 AA;  
SQ Query Match 66.8%; Score 999.5; DB 20; Length 292;  
Best Local Similarity 67.5%; Pred. No. 1.2e-87;  
Matches 199; Conservative 36; Mismatches 55; Indels 5; Gaps 4;  
Qy 1 MNKCLKFTIINTVLVCLSLPNISSKAINNNAKYYGLYISGQYKPSVSFNSFSVKET 60  
Db 1 MNKSKQLIRIFLTLMLSLPNISSK-VNN--EKHSLGLYISGQYKPSVSFNSFSVKET 57  
Qy 61 NVITKNLIALKKDVSIEKTDASVGISNPSNFTPTATVAFQDNSVNFNGTIGYTFACET 119  
Db 58 NFFTQQLVALKKDQDVSIEKTDASVGISNPSNFTPTATVAFQDNTNCSGIGYAFACET 117  
Qy 120 TRVEIGSYEEFDVKNPQGYT-LSDAYRYPALAREMKGNSFTPKKEVSNSIFHTVMRNDG 178  
Db 118 PRIELSYEKEFDVKNPQGYTQKDAYRYFALAREINISLFPQKQEGSGIYHVMKNDG 177  
Qy 179 LSISVIVNVVCDPSLNNLSISPIYICGGAGVDAIEFFDLVHLIKFAYOSKLGIAYSIPSN 238  
Db 178 LSILSNIVNICYDFSLNNLPISPYLCGGMGINAIEFFDLVHLIKFAYOSKAGISYQLLRKI 237







XX PA (RERE-) RES DEV FOUND.  
 XX FI Walker DH, Yu X;  
 XX DR WPI; 2002-066527/09.  
 XX PS Claim 10; Figure 2; 97pp; English.  
 XX CC The invention relates to isolated and purified 28-kDa outer membrane proteins (OMP) from Ehrlichia chaffeensis 28-kDa outer membrane protein, designated P28 useful as a vaccine against Ehrlichia chaffeensis -  
 XX CC Claim 10; Figure 2; 97pp; English.  
 XX CC The invention relates to isolated and purified 28-kDa outer membrane proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins are encoded by a 28kDa outer membrane protein multigene family. P28 proteins are useful as a vaccine against E. chaffeensis. DNA encoding P28 is useful for transfecting a host cell. AAU73400-AAU73420 represent Ehrlichia chaffeensis P28 outer membrane proteins of the invention.  
 XX SQ Sequence 279 AA;  
 Query Match 52.1%; Score 779.5; DB 23; Length 279;  
 Best Local Similarity 54.3%; Pred. No. 1.6e-66;  
 Matches 159; Conservative 39; Mismatches 78; Indels 17; Gaps 4;  
 QY 3 NKLKFTIINTVLVCLLSLPNIISSSKAINNN--AKKYGLIYISQYKPSVSVFSNFSVKET 60  
 DB 2 SKKKFTITGTVLASLLSFLSIESFSAINHNHTGNTSGIYITGQYRPGVSHFSNFSVKET 61  
 QY 61 NVITKNLIALKKVDSDIETKTDAVSGISNPSNFTIPTAVFODNSVNFNGTIGYTFARGT 120  
 DB 62 NVDTITQLVGYKKSASSIDNT-----YSNFQPYVTTFQDNAASTGALGYSPESL 113  
 QY 121 RVEIEGSYEEDVKNPGGYTILSDAYRYFALAREMKNSFTPKKVSNSIFHTVMRNDGLS 180  
 DB 114 RLEEGSYEKFDVKDPKDYSAKDAFRFFALAR---NTST---TVPDAQKVTVMKNGLS 166  
 QY 181 IISVIVNVCYDFSLNNLSISPIYICGAGVDAIEFDVLHIKFAVQSKLGIAYSLSNLSL 240  
 DB 167 VASIMINGCYDLSFNLVSPYICAGIGEDFIEFFDTLHIKLAQYQKLGISYFFPKINV 226  
 QY 241 PASLYYHKVMGNQFNKLVNQHVAELASIPKITSATVATLNIGYFGGEIGARLTF 293  
 DB 227 FAGGYHVRVIGNKFNLVNHHVTFDEPKATSATVATLNVAIFGGEAGVKFTF 279  
 RESULT 15  
 AAU06955  
 ID AAU06955 standard; Protein; 283 AA.  
 XX AC AAU06955;  
 XX DT 05-JUL-1999 (first entry)  
 XX DE E. chaffeensis OMP-1W protein.  
 XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
 XX KW detection; dog.  
 XX OS Ehrlichia chaffeensis.  
 XX PN WO9913720-A1.  
 XX PD 25-MAR-1999.  
 XX PF 18-SEP-1998; 98WO-US19600.  
 XX PR 19-SEP-1997; 97US-0059353.  
 XX PA (OHIS) UNIV OHIO STATE.  
 XX PI Ohashi N, Rikihisa Y;  
 XX

DR WPI; 1999-254290/21.  
 DR N-PSDB; AAU34755.  
 XX Novel outer membrane proteins from Ehrlichia chaffeensis and  
 PT Ehrlichia canis  
 XX XX  
 XX Disclosure; Fig 15B; 55pp; English.  
 XX PS The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 CC in AAU06943-958. The E. canis proteins form part of the P30 family and  
 CC consist of proteins shown in AAU06959-970. The proteins and genes are  
 CC used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX SQ Sequence 283 AA;  
 Query Match 51.3%; Score 768; DB 20; Length 283;  
 Best Local Similarity 54.4%; Pred. No. 2.1e-65;  
 Matches 161; Conservative 41; Mismatches 78; Indels 16; Gaps 6;  
 QY 1 MNKLKFTIINTVLVCLLS-LP--NISSKAINNNAKKYGLIYISQYKPSVSVFSNFSV 57  
 DB 1 MSAKKKLFIIIGSVLVCLVSLPTKSLNLSNNNT-KCTGLYVSGYKPTVSHFSNFSL 59  
 QY 58 KETNVITKNLIALKKVDSDIETKTDAVSGISNPSNFTIPTAVFODNSVNFNGTIGYTF 117  
 DB 60 KETVTDKELLGLLAKDIKSI---TD----ITNKKFNIPYNTKQDNVAFSAAGVYISQ 112  
 QY 118 EGTRVEIEGSYEEDVKNPGGYTILSDAYRYFALAREMKNSFTPKKVSNSIFHTVMRND 177  
 DB 113 DSPRVEVMSYEEDVKNPGVYVSEAFRIYALARGIDN-----LQKYPEITNKVYVKN 167  
 QY 178 GLSIIIVNVCYDFSLNNLSISPIYICGAGVDAIEFDVLHIKFAVQSKLGIAYSLSN 237  
 DB 168 GLSVASIIINGCYDFSLNNLSISPIYICVGFGGDIIEFFSAVSFPKAYQKGVISYPLFSN 227  
 QY 238 ISLFSALYYHKVMGNQFNKLVNQHVAELASIPKITSATVATLNIGYFGGEIGARLTF 293  
 DB 228 MIIFADGYHKVIGNKFNLVNHHVTFDEPKATSATVATLNVAIFGGEAGVKFTF 283  
 RESULT 16  
 AAU73406  
 ID AAU73406 standard; Protein; 283 AA.  
 XX AC AAU73406;  
 XX DT 12-MAR-2002 (first entry)  
 XX DE Ehrlichia chaffeensis outer membrane protein P28-7.  
 XX KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.  
 XX OS Ehrlichia chaffeensis.  
 XX PN WO200183699-A2.  
 XX PD 08-NOV-2001.  
 XX PF 01-MAY-2001; 2001WO-US13997.  
 XX PR 01-MAY-2000; 2000US-201035P.  
 XX PA (RERE-) RES DEV FOUND.  
 XX PI Walker DH, Yu X;  
 XX WPI; 2002-066527/09.  
 XX Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated  
 PT P28 useful as a vaccine against Ehrlichia chaffeensis -  
 XX







Qy	64	TKNLIALKQVDSTETKTDASVGISNPNFTPIPTAVFQDNSVNFNGTIGYTFAGETRV	121
Db	64	V-DLIGLTTDVTYITEHI-----LRDNTKFNTHYIAKFNPNINFSSAIGYSGQGRLE	117
Qy	124	IEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPK-----EKVNSNIFHTVMR	175
Db	118	IESYGDVDVYKNYAVQDVNRIPALVREKNGSNFSPAPHETSQPSDNPKNKSYFTLMK	177
Qy	176	NDGLSIISVIVNVCYDFSLNLSISPIYICGGAGVDAlEFFDVLHLKFAYQSKLGIAYSLP	235
Db	178	NGGVFASVVIINGCYDFSFNNTTISPYVICGVGGDFIEFFEVMHIKFAQSKVGISYPI	237
Qy	236	SNISLFSALYHKMGNOFKNLINQHVLAELASIPKITSATVATLNICYGSGEIGARLTF	293
Db	238	PSITIFADAHYKHVINNKFNHLVKYSYELKNSPTTITSATAKLNIYFGEVGMRFIF	295
RESULT 21			
Qy	AAU73404		
ID	AAU73404	standard; Protein; 295 AA.	
XX	AAU73404;		
XX	12-MAR-2002	(first entry)	
XX	Ehrlichia chaffeensis	outer membrane protein P28-5.	
XX	Ehrlichia;	outer membrane protein; P28; antibiotic; vaccine.	
XX	Ehrlichia chaffeensis.		
XX	WO200183699-A2.		
XX	08-NOV-2001.		
XX	01-MAY-2001;	2001WO-US13997.	
XX	01-MAY-2000;	2000US-201035P.	
XX	(RERE-) RES DEV FOUND.		
XX	Walker DH, Yu X;		
XX	WPI; 2002-066527/09.		
XX	Novel Ehrlichia chaffeensis	28-kDa outer membrane protein, designated	
XX	P28 useful as a vaccine against Ehrlichia chaffeensis		
XX	Claim 10; Figure 2; 97pp; English.		
XX	The invention relates to isolated and purified 28-kDa outer membrane		
XX	proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins		
XX	are encoded by a 28kDa outer membrane protein multigene family. P28		
XX	proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28		
XX	is useful for transfecting a host cell. AAU73400-AAU73420 represent		
XX	Ehrlichia chaffeensis P28 outer membrane proteins of the invention.		
XX	Sequence	295 AA;	
SQ	Query Match	44.9%; Score 671; DB 23; Length 295;	
	Best Local Similarity	46.6%; Pred. No. 4.7e-56;	
	Matches 139; Conservative	50; Mismatches 93; Indels 16; Gaps 4	
Qy	6	KFTIINTVLVCLLSLPNISS--SKAINNNAKKYGLYISGQYKPSVSVFSPNFKETNVI	63
Db	4	KFNFNVLITLFLFLPLKSFYYANNNTITQKVLGIYSGGYKPSIPHFKNFSVEENDKV	63
Qy	64	TKNLIALKQVDSTETKTDASVGISNPNFTPIPTAVFQDNSVNFNGTIGYTFAGETRV	123
Db	64	V-DLIGLTTDVTYITEHI-----LRDNTKFNTHYIAKFNPNINFSSAIGYSGQGRLE	117
Qy	124	IEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPK-----EKVNSNIFHTVMR	175

```
RESULT 23
AAU73403
ID AAU73403 standard; Protein; 272 AA.
XX
AC AAU73403;
XX
DT 12-MAR-2002 (first entry)
XX
DE Ehrlichia chaffeensis outer membrane protein P28-4.
XX
KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.
XX
OS Ehrlichia chaffeensis.
XX
PN WO200183699-A2.
XX
PD 08-NOV-2001.
XX
PF 01-MAY-2001; 2001WO-US13997.
XX
PR 01-MAY-2000; 2000US-201035P.
XX
PA (RERE-) RES DEV FOUND.
XX
PI Walker DH, Yu X;
XX
DR WPI; 2002-066527/09.
XX
PT Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated
PT P28 useful as a vaccine against Ehrlichia chaffeensis -
XX
PS Claim 10; Figure 2; 97pp; English.
XX
CC The invention relates to isolated and purified 28-kDa outer membrane
CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins
CC are encoded by a 28kDa outer membrane protein multigene family. P28
CC proteins are useful as a vaccine against E. chaffeensis. DNA encoding P28
CC is useful for transfecting a host cell. AAU73400-AAU73420 represent
CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.
XX
SQ Sequence 272 AA;
Query Match 34.7%; Score 519; DB 23; Length 272;
Best Local Similarity 39.7%; Pred. No. 1.7e-41;
Matches 118; Conservative 50; Mismatches 97; Indels 32; Gaps 7;
QY 1 MNKLKFTIINTVLVCL--LSLPNISSSKAINNNNAK--KYGLYISGQYKPSVSVSNFS 56
DB 3 MYNKHVCYIVTVITLFLPLLPIESLSALIGNVEKOLKVSSTYVSSQYKPSIFHRNFS 62
QY 57 VKETVITNKLIALKKVDSDIETKTDASVGINSPNFTIPTAVFDQNSVNFNGTIGYTF 116
DB 63 IOESHP-----KKSSEEFK-KIKANLNILKSN---AYNLQPDQNTTSPSGTIGY-F 109
QY 117 AEGTRVELEGSEYEEFDVKNPGGYTLSDAVRYFALAREMKGNSTFTPEKVSNSIFHTVMRN 176
DB 110 SKGLLEAEGCYQEENRVNKNNSLISSNKHSHRIDEN-----VAITTN 154
QY 177 DGLSISIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDVLHIKFAYQSKLGIAYSPLS 236
DB 155 NKLSIASIWNVTCYDISINNTSIVPVLCTGIGEDLVGLFNTIHFKLAYQGVGMSYLNIN 214
QY 237 NISLFASLYHKMGQNFKNLVQHVAELASIPKITSAVATLNIGYFGGEIGARLTF 293
DB 215 NILLFSDIYHYKVMGRFNKLYMQYVADPNISEETIPIILAKLDIGYFGGEIGIREMF 271
RESULT 24
AAU06969
ID AAU06969 standard; Protein; 161 AA.
XX
AC AAU06969;
XX
DT 05-JUL-1999 (first entry)
XX
DE E. canis P30-9 protein.
XX
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
XX
OS Ehrlichia canis.
XX
PN WO9913720-A1.
XX
PD 25-MAR-1999.
XX
PF 18-SEP-1998; 98WO-US19600.
XX
PR 19-SEP-1997; 97US-0059353.
XX
PA (OHIS ) UNIV OHIO STATE.
XX
PI Ohashi N, Rikihisa Y;
XX
DR WPI; 1999-254290/21.
XX
DR N-PSDB; AAX34769.
XX
PT Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
XX
PS Disclosure; Fig 29B; 55pp; English.
XX
CC The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAU06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAU06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
SQ Sequence 161 AA;
Query Match 33.5%; Score 500.5; DB 20; Length 161;
Best Local Similarity 59.3%; Pred. No. 4.8e-40;
Matches 99; Conservative 20; Mismatches 41; Indels 7; Gaps 2;
QY 127 SYEEFDVKNPGGYTLSDAVRYFALAREMKGNSTFTPEKVSNSIFHTVMRNDGLSIVIV 186
DB 2 SYEEFDVKNPGYST-TDSYRYFALARGMDGNIFTOK-----FTVMRNDGLISSVMI 54
QY 187 NVCYDFSILNNLSISPYICGGAGVDAIEFFDVLHIKFAYQSKLGIAYSPLSISLFASLY 246
DB 55 NGCYNVILNDIQAEPYICAGLGGDFIEFFNGFHVKLAVQGVGISYQIFPEVRLFDGYY 114
QY 247 HKVMGNQFKNLVQHVAELASIPKITSAVATLNIGYFGGEIGARLTF 293
DB 115 HKVGNKFNKLVQHVAELASIPKITSAVATLNIGYFGGEIGARLTF 161
RESULT 25
AAU06944
ID AAU06944 standard; Protein; 283 AA.
XX
AC AAU06944;
XX
DT 05-JUL-1999 (first entry)
XX
DE E. chaffeensis OMP-1B protein.
XX
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
XX
OS Ehrlichia chaffeensis.
XX
PN WO9913720-A1.
XX
PD 25-MAR-1999.
```

```

XX PF 18-SEP-1998; 98WO-US19600.
XX PR 19-SEP-1997; 97US-0059353.
XX PA (OHIS ) UNIV OHIO STATE.
XX PI Ohashi N, Rikihisa Y;
XX DR WPI; 1999-254290/21.
XX DR N-PSDB; AAX34744.
XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and
XX PT Ehrlichia canis
XX PS Claim 12; Fig 4B; 55pp; English.
XX CC The invention provides isolated outer membrane proteins (OMP) from
XX CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
XX CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
XX CC in AAU06943-958. The E. canis proteins form part of the p30 family and
XX CC consist of proteins shown in AAU06959-970. The proteins and genes are
XX CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX SQ Sequence 283 AA;

Query Match 33.3%; Score 498; DB 20; Length 283;
Best Local Similarity 36.1%; Pred. No. 1.9e-39;
Matches 112; Conservative 60; Mismatches 94; Indels 44; Gaps 11;

Qy 1 MNKKLFTIINTVLVCLLSL-----PNTSSKAINNNAKKYGLYISQYKPSVSVP 52
Db 1 MNYKKIF--VSSALISLMSILPYQSFADPVTSTNDTGINDSRE--GFYISVKYNPSISHF 55

Qy 53 SNFSVKETNV-----ITKNLIAKQVDSTETKTDASVGINSPNFTIPYTAV-FQNSV 106
Db 56 RKFAEAEAPINGNTSITKKVFLKKQGD-----IAQSANFNRTDPALSFQNNLI 104

Qy 107 N-FNGTIGYTPAETRVEIEGSEYEEFDVKNPGGYTLS--DAYRYFALAREMKGNSFTPK 163
Db 105 SGFSGSIGYAM-DGPRIELEAAVQKFDKPNDDNTNSGDYKYFGLSRE---DAIADKK 160

Qy 164 KVSNSIFHTVWRNDGLSIIISVINVCYDFSLNLSISPIYICGGAGVDIAEFDFVLHKFA 223
Db 161 -----YVVLKNEGITFMSLMVNTCYDITAEGVPFIPYACAGVGADLINVFKDFNLKFS 213

Qy 224 YQSKLGIAYSPLSNISLPSFASLYYHKVMGNQFKNLVQHVAELASIPKITSAVATLNTGYF 283
Db 214 YQKGIGISYPTPEVSFAFIGGYHGVGNFNKIPVITPVVLEGAPOQTTSALVTIDTGYF 273

Qy 284 GGEIGARLTF 293
Db 274 GGEVGVRETF 283

RESULT 26
AAU96106
ID AAU96106 standard; Protein; 283 AA.
XX AC AAU96106;
XX DT 02-JUL-2002 (first entry)
XX DE Ehrlichia chaffeensis OMP-1B.
XX KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;
XX KW antibacterial.
XX OS Ehrlichia chaffeensis.
XX PN WO20022782-A2.
XX PD 21-MAR-2002.

```

```

XX PF 12-SEP-2001; 2001WO-US28759.
XX PR 12-SEP-2000; 2000US-0660587.
XX PA (RERE-) RES DEV FOUND.
XX PI Walker DH, Yu X, McBride JW;
XX DR WPI; 2002-351882/38.
XX PT New recombinant homologous 28 kilodalton immunodominant protein from
XX PT Ehrlichia canis, useful for treating Ehrlichia canis infections -
XX PS Example 3; Figure 3; 106pp; English.
XX CC The invention relates to a recombinant homologous 28 kDa immunodominant
XX CC protein, p28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
XX CC dispersed in a pharmaceutically acceptable carrier, is useful for
XX CC inhibiting E. canis infection in a subject. (I) is useful in the
XX CC development of vaccines and serodiagnostics that are particularly
XX CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
XX CC represent the 28-kDa antigen amino acid sequences of the invention.
XX SQ Sequence 283 AA;

Query Match 33.3%; Score 498; DB 23; Length 283;
Best Local Similarity 36.1%; Pred. No. 1.9e-39;
Matches 112; Conservative 60; Mismatches 94; Indels 44; Gaps 11;

Qy 1 MNKKLFTIINTVLVCLLSL-----PNTSSKAINNNAKKYGLYISQYKPSVSVP 52
Db 1 MNYKKIF--VSSALISLMSILPYQSFADPVTSTNDTGINDSRE--GFYISVKYNPSISHF 55

Qy 53 SNFSVKETNV-----ITKNLIAKQVDSTETKTDASVGINSPNFTIPYTAV-FQNSV 106
Db 56 RKFAEAEAPINGNTSITKKVFLKKQGD-----IAQSANFNRTDPALSFQNNLI 104

Qy 107 N-FNGTIGYTPAETRVEIEGSEYEEFDVKNPGGYTLS--DAYRYFALAREMKGNSFTPK 163
Db 105 SGFSGSIGYAM-DGPRIELEAAVQKFDKPNDDNTNSGDYKYFGLSRE---DAIADKK 160

Qy 164 KVSNSIFHTVWRNDGLSIIISVINVCYDFSLNLSISPIYICGGAGVDIAEFDFVLHKFA 223
Db 161 -----YVVLKNEGITFMSLMVNTCYDITAEGVPFIPYACAGVGADLINVFKDFNLKFS 213

Qy 224 YQSKLGIAYSPLSNISLPSFASLYYHKVMGNQFKNLVQHVAELASIPKITSAVATLNTGYF 283
Db 214 YQKGIGISYPTPEVSFAFIGGYHGVGNFNKIPVITPVVLEGAPOQTTSALVTIDTGYF 273

Qy 284 GGEIGARLTF 293
Db 274 GGEVGVRETF 283

RESULT 27
AAU73413
ID AAU73413 standard; Protein; 283 AA.
XX AC AAU73413;
XX DT 12-MAR-2002 (first entry)
XX DE Ehrlichia chaffeensis outer membrane protein p28-14.
XX KW Ehrlichia; outer membrane protein; p28; antibiotic; vaccine.
XX KW Ehrlichia chaffeensis.
XX OS Ehrlichia chaffeensis.
XX PN WO200183699-A2.
XX PD 08-NOV-2001.

```

```
PF 01-MAY-2001; 2001WO-US13997.
XX
PR 01-MAY-2000; 2000US-201035P.
XX
XX (RERE-) RES DEV FOUND.
XX
PI Walker DH, Yu X;
XX
XX WPI; 2002-066527/09.
DR
XX
XX Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated
PT P28 useful as a vaccine against Ehrlichia chaffeensis
XX
XX Disclosure; Figure 2; 97pp; English.
XX
XX The invention relates to isolated and purified 28-kDa outer membrane
CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins
CC are encoded by a 28kDa outer membrane protein multigene family. P28
CC proteins are useful as a vaccine against E. chaffeensis. DNA encoding P28
CC is useful for transfecting a host cell. AAU73400-AAU73420 represent
CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.
XX
XX Sequence 283 AA;
SQ
    Query Match      33.3%; Score 498; DB 23; Length 283;
    Best Local Similarity 36.1%; Pred. No. 1.9e-39;
    Matches 112; Conservative 60; Mismatches 94; Indels 44; Gaps 11;
QY 1 MNKLKFTIINTVLVCLLSL-----PNISSKAINNNAKYYGLYISGYKPSVSVF 52
DB 1 MNYKKIF--VSSALISLSMLPYQSFADPVTSDTNDTGINDSRE---GFYISVKYNSISHF 55
QY 53 SNFSVKETNV-----ITKNLIAKXKDVDSIETKTDASVGSINPSNFTTIPYTA-V-FQDMSV 106
DB 56 RKFSAEAEAPINGNTSITKKVFGKDKGD-----IAQSANFNRTDPALEFQNNLI 104
QY 107 N-FNCTIGVTFAGTRVEGSEYEDVKNGPGGYTLSDA--VRYFALAREMKGNSFTPK 163
DB 105 SGFSGSIGYAM-DGPRIELEAAAYQFDPKNDPNDNTSGDYKYFGLSRE---DAIADKK 160
QY 164 KVSNSIFHTVMENDGLSIISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDVLHIFKA 223
DB 161 -----YVVLKNEGITFMSLVNTCYDITAEGVFPPIYACAGVGADLINVFKDFNLKFS 213
QY 224 YQSLGSIAYSLPSNLSLFASLYHYKMGNOFKNLNVQHVLAELASPKITSATVATNLNIGYF 283
DB 214 YQKIGISYPITPEVSAFTGGYHYGVGNKFKIPVITPVVLEGAPOTTTSALVTIDTGYF 273
QY 284 GGEIGARLTF 293
DB 274 GGEVGVRTTF 283
RESULT 28
AAU96116
ID AAU96116 standard; Protein; 280 AA.
XX
XX AAU96116;
AC
XX
XX 02-JUL-2002 (first entry)
DT
XX
XX Ehrlichia canis p28-2.
DE
XX
XX Ehrlichia canis infection; vaccine; serodiagnostic; p28;
KW
KW antibacterial.
XX
XX Ehrlichia canis.
OS
XX
XX WO200222782-A2.
PN
XX
XX 21-MAR-2002.
PD
XX
XX 12-SEP-2001; 2001WO-US28759.
PF
```

```
XX
PR 12-SEP-2000; 2000US-0660587.
XX
XX (RERE-) RES DEV FOUND.
XX
XX Walker DH, Yu X, McBride JW;
XX
XX WPI; 2002-351882/38.
DR
XX
XX N-PSDB; ABK68876.
XX
XX New recombinant homologous 28 kilodalton immunodominant protein from
PT Ehrlichia canis, useful for treating Ehrlichia canis infections
XX
XX Claim 16; Figure 14; 106pp; English.
XX
XX The invention relates to a recombinant homologous 28 kDa immunodominant
CC protein, p28, (1), of Ehrlichia canis. (1), a 28-kDa antigen preferably
CC dispersed in a pharmaceutically acceptable carrier, is useful for
CC inhibiting E. canis infection in a subject. (1) is useful in the
CC development of vaccines and serodiagnostics that are particularly
CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
CC represent the 28-kDa antigen amino acid sequences of the invention.
XX
XX Sequence 280 AA;
SQ
    Query Match      32.8%; Score 490.5; DB 23; Length 280;
    Best Local Similarity 37.1%; Pred. No. 9.9e-39;
    Matches 112; Conservative 60; Mismatches 91; Indels 39; Gaps 11;
QY 6 KFTIINTVLVCLLS-LPNISSSKAI-----NNNAKKYYGLYISGYKPSVSVFSVKE- 59
DB 4 KKLIVRSALISLSMLPYQSFADPVGSRTRNDNKE---GFYISAKYNSISHFRKFSAEET 60
QY 60 ---TNVITKNLIAKXKDVDSIETKTDASVGSINPSNFTTIPYTA-V-FQDMSV-N-FNGTIG 113
DB 61 PINGTNSLITKKVFGKDKGD-----ITKKDDFTRVAPGIDFQNNLISGFSIG 109
QY 114 YTFAGTRVEGSEYEDVKNGPGGYTLSDA--VRYFALAREMKGNSFTPKKYSNIFH 171
DB 110 YSM-DGPRIELEAAAYQFDPKNDPNDNTDNGEYKYHFALSR-----KDAMEDQOY- 158
QY 172 TWMENDGLSIISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDVLHIFKAYQSLGIA 231
DB 159 VVLKNDGITFMSLVNTCYDITAEGVSVFPVACAGIGADLITIFKDLNLKFAVQKIGIS 218
QY 232 YSLPSNLSLFASLYHYKMGNOFKNLNVQHVLAELASPKITSATVATNLNIGYFGGEIGARL 291
DB 219 YPITPEVSAFTGGYHYGVGNKFKIPVITPVVLENDAPQTTTSASVTLDDVGYFGGEIGMRF 278
QY 292 TF 293
DB 279 TF 280
RESULT 29
AAU96948
ID AAU96948 standard; Protein; 280 AA.
XX
XX AAU96948;
AC
XX
XX 05-JUL-1999 (first entry)
DT
XX
XX E. chaffeensis OMP-1F protein.
DE
XX
XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog.
XX
XX Ehrlichia chaffeensis.
OS
XX
XX WO9913720-A1.
PN
XX
XX 25-MAR-1999.
PD
XX
```

PF 18-SEP-1998; 98WO-US19600.  
 XX  
 PR 19-SEP-1997; 97US-0059353.  
 XX  
 PA (OHIS ) UNIV OHIO STATE.  
 XX  
 PI Ohashi N, Rikihisa Y;  
 XX  
 DR WPI; 1999-254290/21.  
 XX  
 DR N-PSDB; AAX34748.  
 XX

XX Novel outer membrane proteins from Ehrlichia chaffeensis and  
 PT Ehrlichia canis  
 XX

PS Claim 16; Fig 8B; 55pp; English.

XX The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 CC in AA06943-958. The E. canis proteins form part of the p30 family and  
 CC consist of proteins shown in AA06959-970. The proteins and genes are  
 CC used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX

SQ Sequence 280 AA;

Query Match 31.5%; Score 470.5; DB 20; Length 280;  
 Best Local Similarity 37.5%; Pred. No. 8.3e-37;  
 Matches 115; Conservative 57; Mismatches 94; Indels 41; Gaps 14;

Qy 1 MNKKLFTIINTVLVCLLS-LPNISSKAINNNAKYGLYISQYKPSVSNFSVKE 59  
 Db 1 MNCK-KP-FITTLVLSLMSFLPGISFSDAVQND-NVGGNFYISGKYPSVSHFGVSAKQ 57  
 Qy 60 TNVITKNLIALKKDVSIEETKTDAVGISNPSN-FTIP-YTAVFDNS-VNFGTIGYTF 116  
 Db 58 ERNTTGVGLKQWDG-----STISKSPENTFVNYSFKYENPFPGAGAGY-L 110  
 Qy 117 ABGTRVEIGSYEEFDVKNPGGYTLSDAYRYFALAREMKNSTPPEKVSNS----IPHT 172  
 Db 111 MNGPRIELMSYETFDVKNQGNKYKNDAAHYALTHNSGG-----KLSNAGDKFVF-- 161  
 Qy 173 VMNDGLSIISVIVNVCYDFSLNNLSISPYICGAGVDAIEFFDLVHIKAYQSKLGIAI 232  
 Db 162 -LKNEGLDISLMLNACYDVISEGIPFPYICAGVGTDLISMFEAINPKISYQKGLGSLY 220  
 Qy 233 SLPSNISLFASLYYHKVMGNQKMLNVQHVLAELASIPKITS-----AVATLNIGYFGGE 286  
 Db 221 SISPEASVFGGHFKHVGNEFRDIP-----AMIPSTLTGNGHFTITVLSVCHFGVE 273  
 Qy 287 IGARLTF 293  
 Db 274 LGGRFNF 280

RESULT 30  
 AAU96110  
 ID AAU96110 standard; Protein; 280 AA.  
 XX  
 AC AAU96110;

DT 02-JUL-2002 (first entry)

XX Ehrlichia chaffeensis OMP-1P.

XX Ehrlichia canis infection; vaccine; serodiagnostic; p28;  
 KW antibacterial.  
 XX

OS Ehrlichia chaffeensis.

XX WO20022782-A2.

XX 21-MAR-2002.

XX

PF 12-SEP-2001; 2001WO-US28759.  
 XX  
 PR 12-SEP-2000; 2000US-0660587.  
 XX  
 PA (RERE-) RES DEV FOUND.  
 XX  
 PI Walker DH, Yu X, McBride JW;  
 XX  
 DR WPI; 2002-351882/38.  
 XX

XX New recombinant homologous 28 kilodalton immunodominant protein from  
 PT Ehrlichia canis, useful for treating Ehrlichia canis infections  
 XX

PS Example 3; Figure 3; 106pp; English.

XX The invention relates to a recombinant homologous 28 kDa immunodominant  
 CC protein, p28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably  
 CC dispersed in a pharmaceutically acceptable carrier, is useful for  
 CC inhibiting E. canis infection in a subject. (I) is useful in the  
 CC development of vaccines and serodiagnostics that are particularly  
 CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118  
 CC represent the 28-kDa antigen amino acid sequences of the invention.  
 XX

SQ Sequence 280 AA;

Query Match 31.5%; Score 470.5; DB 23; Length 280;  
 Best Local Similarity 37.5%; Pred. No. 8.3e-37;  
 Matches 115; Conservative 57; Mismatches 94; Indels 41; Gaps 14;

Qy 1 MNKKLFTIINTVLVCLLS-LPNISSKAINNNAKYGLYISQYKPSVSNFSVKE 59  
 Db 1 MNCK-KP-FITTLVLSLMSFLPGISFSDAVQND-NVGGNFYISGKYPSVSHFGVSAKQ 57  
 Qy 60 TNVITKNLIALKKDVSIEETKTDAVGISNPSN-FTIP-YTAVFDNS-VNFGTIGYTF 116  
 Db 58 ERNTTGVGLKQWDG-----STISKSPENTFVNYSFKYENPFPGAGAGY-L 110  
 Qy 117 ABGTRVEIGSYEEFDVKNPGGYTLSDAYRYFALAREMKNSTPPEKVSNS----IPHT 172  
 Db 111 MNGPRIELMSYETFDVKNQGNKYKNDAAHYALTHNSGG-----KLSNAGDKFVF-- 161  
 Qy 173 VMNDGLSIISVIVNVCYDFSLNNLSISPYICGAGVDAIEFFDLVHIKAYQSKLGIAI 232  
 Db 162 -LKNEGLDISLMLNACYDVISEGIPFPYICAGVGTDLISMFEAINPKISYQKGLGSLY 220  
 Qy 233 SLPSNISLFASLYYHKVMGNQKMLNVQHVLAELASIPKITS-----AVATLNIGYFGGE 286  
 Db 221 SISPEASVFGGHFKHVGNEFRDIP-----AMIPSTLTGNGHFTITVLSVCHFGVE 273  
 Qy 287 IGARLTF 293  
 Db 274 LGGRFNF 280

RESULT 31  
 AAU06959  
 ID AAU06959 standard; Protein; 288 AA.  
 XX  
 AC AAU06959;

DT 05-JUL-1999 (first entry)

XX E. canis p30 protein.

XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; p30;  
 KW detection; dog.  
 XX

OS Ehrlichia canis.

XX WO9913720-A1.

XX 25-MAR-1999.

XX

PF 18-SEP-1998; 98WO-US19600.  
 XX  
 PR 19-SEP-1997; 97US-0059353.  
 XX  
 PA (OHIS ) UNIV OHIO STATE.  
 XX  
 PI Ohashi N, Rikihisa Y;  
 XX  
 DR WPI; 1999-254290/21.  
 XX  
 DR N-PSDB; AAX34759.  
 PT  
 PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
 XX Ehrlichia canis  
 PS Disclosure; Fig 19B; 55pp; English.  
 XX  
 CC The invention provides isolated outer membrane proteins (OMP) from  
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and  
 CC consist of proteins shown in AAY06959-970. The proteins and genes are  
 CC used to detect E. chaffeensis in patients and E. canis in dogs.  
 XX  
 SQ Sequence 288 AA;  
 Query Match 31.5%; Score 470.5; DB 20; Length 288;  
 Best Local Similarity 38.3%; Pred. No. 8.6e-37;  
 Matches 120; Conservative 51; Mismatches 97; Indels 45; Gaps 12;  
 QY 1 MNKKLFTIINTVLVCLLS-LPNISSKA-----INNNAKYYGLYISGYKPSVSFVSN 54  
 DB 1 MNCKRFP--IASALISLMSFLPSVSFSESIHEDNING-----FYISAKYMPSSHFGV 52  
 QY 55 FSVKTEVITKNLIAALKKDVSIEITKTDASVGSNPSN-FTIP-YTAVFQDMS-VNENGT 111  
 DB 53 FSVKEKNTTGVGLKQDWDGATIKDASSHTIDPSTIFSISNYSFKYNNPFLGPAGA 112  
 QY 112 IGYTAEGRVIEGSEYEFVKNVGGYTLSDAYRYFALAREMKNSFTPKKVSNSIFH 171  
 DB 113 IGYSMG-GPRVEFVSIEYFDVKNQGNYSYKNDAAHYKALSRHTGG---MPQAGHQKFP-- 166  
 QY 172 TWRNDGLSIIISVIVNVCYDFSLNLSISPYICGGAGVDAIEFFDVLHFKPAYQSKLGIA 231  
 DB 167 VFLKNEGLDLSMLNACVDITIDSMPPSYICAGISGLVSMFTTTPKISYQKLGVS 226  
 QY 232 YSLPSNISLFLASYHYHKVGNQFNKLNVOHVAELASIPKITSAVA-----TLNI 280  
 DB 227 YSISPEASVFGCHFRVIGNEFKD-----IPAITPAGATEIKGTQFTTITLNI 275  
 QY 281 GYFGGEIGARLTF 293  
 DB 276 CHFGLELGRFTF 288  
 RESULT 32  
 AAW51094  
 ID AAW51094 standard; Protein; 280 AA.  
 XX  
 AC AAW51094;  
 XX  
 DT 14-SEP-1998 (first entry)  
 XX  
 DE Ehrlichia chaffeensis VSA4 protein.  
 XX  
 KW MAP1 homologue; variable surface antigen; VSA4; rickettsia;  
 KW DNA vaccine.  
 XX  
 OS Ehrlichia chaffeensis.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..25  
 FT /note= "putative signal peptide"  
 XX

PN WO9816554-A1.  
 XX  
 PD 23-APR-1998.  
 XX  
 PF 17-OCT-1997; 97WO-US19044.  
 XX  
 PR 17-OCT-1996; 96US-0733230.  
 XX  
 PA (UYFL ) UNIV FLORIDA.  
 XX  
 PI Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;  
 PI Nyika A, Rurangirwa FR;  
 XX  
 DR WPI; 1998-251232/22.  
 DR N-PSDB; AAV07179.  
 XX  
 CC Composition containing nucleic acid encoding rickettsial antigen -  
 CC useful for, e.g. stimulating protective immune response in humans or  
 CC animals  
 XX  
 PS Claim 3; Fig 2B; 39pp; English.  
 XX  
 CC This is the full-length variable surface antigen VSA4 protein of  
 CC Ehrlichia chaffeensis. Its amino acid sequence was deduced from a  
 CC partial open reading frame (ORF4) of a genomic locus (see AAV07179)  
 CC of E. chaffeensis that was obtained on the basis of homology to the  
 CC major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium.  
 CC This genomic locus included 5 ORFs encoding similar, but  
 CC non-identical proteins (see AAW51091-95). A claimed composition  
 CC comprises a nucleic acid (see AAV07176-82) encoding a polypeptide  
 CC (see AAW51088-99) that elicits a protective immune response against a  
 CC rickettsial pathogen. The nucleic acid is used, in human or  
 CC veterinary medicine, in vaccines to protect against Rickettsia,  
 CC Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic  
 CC polypeptides can be used diagnostically to detect antibodies  
 CC associated with Ehrlichia infection (claimed).  
 XX  
 SQ Sequence 280 AA;  
 Query Match 31.3%; Score 468.5; DB 19; Length 280;  
 Best Local Similarity 37.5%; Pred. No. 1.3e-36;  
 Matches 115; Conservative 57; Mismatches 94; Indels 41; Gaps 14;  
 QY 1 MNKKLFTIINTVLVCLLS-LPNISSSKAINNAKYYGLYISGYKPSVSFVSFVSK 59  
 DB 1 MNCK-KP-FITTLVLSMSELPGISFSDAVQND-NVGGNFYISGYKPSVSHFGVFSKQ 57  
 QY 60 TNVITKNLIAALKKDVSIEITKTDASVGSNPSN-FTIP-YTAVFQDMS-VNENGTIGYTF 116  
 DB 58 ERNTTIGVGLKQDWDG-----STISKNSPENTFNPNYSFKYNNPFLGFAGAVGY-L 110  
 QY 117 AEGTRVIEGSEYEFVKNVGGYTLSDAYRYFALAREMKNSFTPKKVSNS-----IPHT 172  
 DB 111 MNGPRIELEMSEYETFDVKNQGNYSYKNDAAHYKALYTHNSGG-----KLSNAGDKFVP-- 161  
 QY 173 VWRNDGLSIIISVIVNVCYDFSLNLSISPYICGGAGVDAIEFFDVLHFKPAYQSKLGIA 232  
 DB 162 -LKNEGLDLSMLNACVDVISEGIPSPYICAGVTDLSIMFPAINKISYQKLGUSY 220  
 QY 233 SLPSNISLFLASYHYHKVGNQFNKLNVOHVAELASIPKITS-----AVATLNICYFGGE 286  
 DB 221 SISPEASVFGCHFRVIGNEFRDIP-----AWIPSTSLTGNHFTITLSVCHFGVE 273  
 QY 287 IGARLTF 293  
 DB 274 LGGRFNF 280  
 RESULT 33  
 AAB36188  
 ID AAB36188 standard; Protein; 280 AA.  
 XX  
 AC AAB36188;

```
XX DT 02-MAR-2001 (first entry)
XX DE Ehrlichia chaffeensis partial VSA4.
XX KW Ehrlichia chaffeensis; VSA4; variable surface antigen 4; MAP1;
XX KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
XX KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hwo1f1; 18hwo1f1;
XX KW 39dor1f3.
XX OS Ehrlichia chaffeensis.
XX XX WO200005063-A2.
XX PN 02-NOV-2000.
XX PD 21-APR-2000; 2000WO-US10886.
XX PF 22-APR-1999; 99US-0130725.
XX PR (UYFL ) UNIV FLORIDA.
XX PA Barbet AF, Bowie MV, Ganta RR, Burrigide MJ, Mahan SM, McGuire TC;
XX PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Allenman AR;
XX XX WPI: 2000-679675/66.
XX DR N-PSDB; AAC68705.
XX XX New polynucleotides useful as DNA vaccines for conferring immunity to
XX PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
XX PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
XX PT -
XX PS Claim 3; Page 45-46; 63pp; English.
XX CC The present sequence shows a high degree of similarity to the major
XX CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
XX CC used in a vaccines to protect animals or humans against rickettsial
XX CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
XX CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
XX CC protective against the rickettsial pathogen. The nucleic acid vaccines
XX CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
XX CC Cowdria ruminantium genes designated map 2, lhworf3, 4hwo1f1, 18hwo1f1
XX CC and 39dor1f3 may be used in therapeutic and diagnostic applications. The
XX CC polypeptides are useful for detecting antibodies associated with
XX CC infection by a rickettsial pathogen whilst the polynucleotides may be
XX CC used to detect the presence of rickettsial nucleic acids.
XX SQ Sequence 280 AA;
XX Query Match 31.3%; Score 468.5; DB 21; Length 280;
XX Best Local Similarity 37.5%; Pred. No. 1.3e-36;
XX Matches 115; Conservative 57; Mismatches 94; Indels 41; Gaps 14;
QY 1 MNKKLKEITINTVLVCLLS-LPNISSKAINNAKKYGLYISQYKPSVSVFNSVKE 59
DB 1 MNCK-KF-FITTLVLSMLSPGIFSDDAVQND-NVGGNFYISGKYPSVSHFGVSAKQ 57
QY 60 TNVITKNLIALKDDVDSIETKTDAVGSINPSN-FTIP-YTAVFDNS-VNFNCTIGVTF 116
DB 58 ERNTTIGVGLKQWDG-----STISKNSPENTFNVNFKYNNPFLGAGAGVY-L 110
QY 117 AEGTRVEIEGSYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNS-----IFHT 172
DB 111 MNGPRIELMSYETFQVKNQGNKNDKHYALTHNSGG-----KLSNAGDKFV-- 161
QY 173 VMNDGLSIISVIVNVCYDLSNLNLSIPYICGAGVDATBEFVLHFKAYQSKLGIAY 232
DB 162 -LKNEGLDLSMLNACDYISEGIPSPYICAGVGTDLISMFEAINPKISYQKGLGSLY 220
QY 233 SLPSNISLFSASYHVKMGQFNKLNQHVLAELASIPKITS-----AVATLNIGYEGE 286
DB 221 SISPEASVFGGHPKHVIGNEFRDIP-----AMIPSTSTLTGNHHTITVLSVCHFGVE 273
```

```
QY 287 IGARLTF 293
DB 274 LGGRFNF 280

RESULT 34
AAU04198
ID AAU04198 standard; Protein; 280 AA.
XX AC AAU04198;
XX DT 23-OCT-2001 (first entry)
XX DE Variable surface antigen 4 (VSA4) from Ehrlichia chaffeensis.
XX KW Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
XX KW infection; heartwater; diagnostic; variable surface antigen; VSA.
XX OS Ehrlichia chaffeensis.
XX PN US6251872-B1.
XX PD 26-JUN-2001.
XX PF 17-OCT-1997; 97US-0953326.
XX PR 17-OCT-1996; 96US-0733230.
XX PA (UYFL ) UNIV FLORIDA.
XX PI Barbet AF, Ganta RR, McGuire TC, Burrigide MJ, Nyika A;
XX PI Rurangirwa FR, Mahan SM, Bowie MV, Allenman AR;
XX XX WPI: 2001-424487/45.
XX DR N-PSDB; AAS07578.
XX XX New MAP2 genes and polypeptides useful as vaccines for conferring
XX PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
XX PT as molecular markers in nucleic acid analysis procedures -
XX XX Example 3; Fig 2A-2B; 30pp; English.
XX CC The sequence represents the amino acid sequence of variable surface
XX CC antigen 4 (VSA4) isolated from Ehrlichia chaffeensis, which
XX CC has similarity to major antigen protein (MAP). The MAP polynucleotides
XX CC and polypeptides are useful as vaccines for conferring immunity to
XX CC rickettsia infection, including Cowdria ruminantium causing heartwater.
XX CC The MAP polynucleotides may be used as molecular markers in nucleic acid
XX CC analysis procedures, and to produce the MAP polypeptides, which may
XX CC be used to raise antibodies that are reactive with the polypeptides.
XX CC The nucleic acids may further be used as probes to identify
XX CC complementary sequences within other nucleic acid molecules or genomes,
XX CC where such probes can be applied to identify or distinguish infectious
XX CC strains of organisms in diagnostic procedures or in rickettsial
XX CC research where identification of particular organisms or strains is
XX CC needed.
XX SQ Sequence 280 AA;
XX Query Match 31.3%; Score 468.5; DB 22; Length 280;
XX Best Local Similarity 37.5%; Pred. No. 1.3e-36;
XX Matches 115; Conservative 57; Mismatches 94; Indels 41; Gaps 14;
QY 1 MNKKLKEITINTVLVCLLS-LPNISSKAINNAKKYGLYISQYKPSVSVFNSVKE 59
DB 1 MNCK-KF-FITTLVLSMLSPGIFSDDAVQND-NVGGNFYISGKYPSVSHFGVSAKQ 57
QY 60 TNVITKNLIALKDDVDSIETKTDAVGSINPSN-FTIP-YTAVFDNS-VNFNCTIGVTF 116
DB 58 ERNTTIGVGLKQWDG-----STISKNSPENTFNVNFKYNNPFLGAGAGVY-L 110
QY 117 AEGTRVEIEGSYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNS-----IFHT 172
```

```
Db 111 MNGPRIEMSEYETFDVKNQNNYKNDHKKYALTHNSGG-----KLSNAGDKFVP-- 161
Qy 173 VMRDGLSIISVIVNVCYDFSLNNLSISPYICGAGVDAIEFFDVLHIKFAVQSKLGIAV 232
Db 162 -LKNEGLLDISLMLNACYDVISEGIPSPYICAGVGTDLISMFEAINPKISYQKGLGUSY 220
Qy 233 SLPSNISLASFALSYHKVMGNQFNKLNVOHVAELASIPKITS-----AVATLNTGYFGGE 286
Db 221 SISPEASVFGGHHKVGNEFRDIP-----AMIPSTLTGNHFTIVTLSVCHFGVE 273
Qy 287 IGARLTF 293
Db 274 LGGRFNF 280

RESULT 35
AAU73417
ID AAU73417 standard; Protein; 280 AA.
XX AC AAU73417;
XX DT 12-MAR-2002 (first entry)
XX DE Ehrlichia chaffeensis outer membrane protein P28-18.
XX KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.
XX OS Ehrlichia chaffeensis.
XX PN WO200183699-A2.
XX PD 08-NOV-2001.
XX PF 01-MAY-2001; 2001WO-US13997.
XX PR 01-MAY-2000; 2000US-201035P.
XX PA (RERE-) RES DEV FOUND.
XX PI Walker DH, Yu X;
XX WPI; 2002-066527/09.
XX PT Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated
XX P28 useful as a vaccine against Ehrlichia chaffeensis
XX PS Disclosure; Figure 2; 97pp; English.
XX CC The invention relates to isolated and purified 28-kDa outer membrane
XX proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins
XX are encoded by a 28kDa outer membrane protein multigene family. P28
XX proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28
XX is useful for transfecting a host cell. AAU73400-AAU73420 represent
XX Ehrlichia chaffeensis P28 outer membrane proteins of the invention.
XX SQ Sequence 280 AA;

Query Match 31.3%; Score 468.5; DB 23; Length 280;
Best Local Similarity 37.5%; Pred. No. 1.3e-36;
Matches 115; Conservative 57; Mismatches 94; Indels 41; Gaps 14;
Qy 1 MNKCLKFTIINTVLVCLLS-LPNISSKAINNNAKYYGLYISQYKPSVSVFNSFSVKE 59
Db 1 MNCK-KP-FITTLVSLMSFLGIFSDDAVQND-NVGGNFYISGKYVPSVSHFGVSAKQ 57
Qy 60 TNVITKNLIALKKDVSIEFTKTDASVGISNPSN-FTIP-YTAVFDONS-VNENGITGYTF 116
Db 58 ERNTTIGVGLKQWDG-----STISKNSPENTFNPNYSKYENNPFGLGPAVGY-L 110
Qy 117 AEGTRVEIGSEYEEFDVKNPGGYTTLSDAYRYFALAREMKGNSFTPKKVNS-----IFHT 172
Db 111 MNGPRIEMSEYETFDVKNQNNYKNDHKKYALTHNSGG-----KLSNAGDKFVP-- 161
```

```
Qy 173 VMRDGLSIISVIVNVCYDFSLNNLSISPYICGAGVDAIEFFDVLHIKFAVQSKLGIAV 232
Db 162 -LKNEGLLDISLMLNACYDVISEGIPSPYICAGVGTDLISMFEAINPKISYQKGLGUSY 220
Qy 233 SLPSNISLASFALSYHKVMGNQFNKLNVOHVAELASIPKITS-----AVATLNTGYFGGE 286
Db 221 SISPEASVFGGHHKVGNEFRDIP-----AMIPSTLTGNHFTIVTLSVCHFGVE 273
Qy 287 IGARLTF 293
Db 274 LGGRFNF 280

RESULT 36
AAU73419
ID AAU73419 standard; Protein; 271 AA.
XX AC AAU73419;
XX DT 12-MAR-2002 (first entry)
XX DE Ehrlichia chaffeensis outer membrane protein P28-20.
XX KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.
XX OS Ehrlichia chaffeensis.
XX PN WO200183699-A2.
XX PD 08-NOV-2001.
XX PF 01-MAY-2001; 2001WO-US13997.
XX PR 01-MAY-2000; 2000US-201035P.
XX PA (RERE-) RES DEV FOUND.
XX PI Walker DH, Yu X;
XX WPI; 2002-066527/09.
XX PT Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated
XX P28 useful as a vaccine against Ehrlichia chaffeensis
XX PS Claim 10; Figure 2; 97pp; English.
XX CC The invention relates to isolated and purified 28-kDa outer membrane
XX proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins
XX are encoded by a 28kDa outer membrane protein multigene family. P28
XX proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28
XX is useful for transfecting a host cell. AAU73400-AAU73420 represent
XX Ehrlichia chaffeensis P28 outer membrane proteins of the invention.
XX SQ Sequence 271 AA;

Query Match 31.0%; Score 464; DB 23; Length 271;
Best Local Similarity 38.3%; Pred. No. 3.3e-36;
Matches 115; Conservative 49; Mismatches 98; Indels 38; Gaps 12;
Qy 1 MNKCLKFTIINTVLVCLLS-LPNISSKAINNNAKYYGLYISQYKPSVSVFNSFSVKE 59
Db 1 MNCK-KP-VVGVGVALATLLSFLPDNSFSDANVPEGRK--GFYVGTQYKVGVPNFSNFAEE 56
Qy 60 T-NVITKNLIALKKDVSIEFTKTDASVGISNPSNFTIPYTAVFQONS-VNNGTIGYTF 118
Db 57 TLPGLTKSIFALGDKSS-----ISDHAGFTQAYNPTYASNFAFGVGVIGY-YVN 105
Qy 119 GTRVEIGSEYEEFDVKN---PGGYTSLDAYRYFALARE--MKGNSFTPKKVNSIFHTV 173
Db 106 DFRVEFEGAYENFEPERQWYPEG---GESHKFFALSRESTVQDNKF-----IV 150
Qy 174 MRNDGLSIISVIVNVCYDFSLNNLSISPYICGAGVDAIEFFDVLHIKFAVQSKLGIAV 233
```



```

Db 151 LENDGVIDKSLNVNFCYDIANGSIPLAPYMCAGVADYIKPLGSLPKFSQVRFQVNP 210
Qy 234 LPSNISLFPASLYYHKVMGNQFNLVN-OHVAELASIPKITSAVATLNIYFGGIGARLT 292
Db 211 VSVNMLFGGGYHKVGNRRYERVEIAVHPATLNVKPTTSASATLDDYFGVEGMRFT 270

RESULT 37
AAW51088
ID AAW51088 standard; Protein; 287 AA.
AC AAW51088;
XX
DT 14-SEP-1998 (first entry)
DE Cowdria ruminantium major antigenic protein 1 (MAP1).
KW MAP1 gene; major antigenic protein 1; rickettsia; heartworm;
KW DNA vaccine.
XX
OS Cowdria ruminantium.
XX
FN WO9816554-A1.
XX
PD 23-APR-1998.
XX
PF 17-OCT-1997; 97WO-US19044.
XX
PR 17-OCT-1996; 96US-0733230.
XX
PA (UYFL ) UNIV FLORIDA.
XX
PI Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;
PI Nyika A, Rurangirwa FR;
XX
WPI; 1998-251232/22.
DR N-PSDB; AAV07176.
XX
Composition containing nucleic acid encoding rickettsial antigen -
useful for, e.g. stimulating protective immune response in humans or
animals
XX
Claim 3; Page 15-16; 39pp; English.
XX
This polypeptide comprises the major antigen protein 1 gene (MAP1)
of Cowdria ruminantium, the causative agent of heartwater in
domestic ruminants. It is encoded by the MAP1 gene (see AAV07176).
A claimed composition comprises a nucleic acid (see AAV07176-82)
encoding a polypeptide (see AAW51088-99) that elicits a protective
immune response against a rickettsial pathogen. The nucleic acid
is used, in human or veterinary medicine, in vaccines to protect
against Rickettsia, Ehrlichia, Anaplasma and Cowdria species. The
nucleic acid does not replicate in the host but remains episomal
and capable of expressing polypeptide for at least 19 mch.
XX
Sequence 287 AA;
Query Match 30.7%; Score 459; DB 19; Length 287;
Best Local Similarity 36.1%; Pred. No. 1.1e-35;
Matches 110; Conservative 65; Mismatches 100; Indels 30; Gaps 10;
Qy 1 MNKKLFTIINTVLVCLLS-LPNISSKAINNNAKYGYLYISQYKPSVSNFSVYKE 59
Db 1 MNCKKIF--ITSTLSLVSLPGVSPDSFSDVIOEDSNPAGSVYISAKYMPYASHFGKMSIKE 58
Qy 60 TNVTKNLIALKKVDSEIETKDVASGISNPSNT-IPYTAVFQDNS-VNPNGTIGYTFEA 117
Db 59 DSKNTQTVFLGLKQWDGKVTSPDSS--NTNSTIFTEKDYSPRYENPNFLGPAIGYSM- 115
Qy 118 EGTVEIEGSEVEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKVKSNIHFYVRND 177
Db 116 NGRPIEFVSEYETFDVKNLGNKYNNAHMYCAL--DTAAQNSTNGAGLTTSV---MVKNE 170
```

```

Qy 178 GLSTISIVNVVYDFSNLNNLSISPYICGAGVDAIEFFDLVHLHKFAYQSKLGIAYSUPSN 237
Db 171 NLATNISLMLNACYDILMDGIPVSPYVCAGIGTDLVSVINATNPKLSYQGLGISYSINSE 230
Qy 238 ISLFPASLYYHKVMGNQFNLVNQHVAELASIPKITS-----AVATLNIYFGGIEIG 288
Db 231 ASIFIGGHRVIGNEPK-----DIATLKIFTSKTGISNPGFASATLDVCHFGIEIG 282
Qy 289 ARLTF 293
Db 283 GRFVF 287

RESULT 38
AAB36182
ID AAB36182 standard; Protein; 287 AA.
AC AAB36182;
XX
DT 02-MAR-2001 (first entry)
DE Cowdria ruminantium MAP1.
XX
KW Cowdria ruminantium; MAP1; major antigenic protein 1; antirickettsial;
KW vaccine; gene therapy; Rickettsia; Ehrlichia; Anaplasma; map2; lhworf3;
KW 4hworf1; 18hworf1; 3gdorf3.
XX
OS Cowdria ruminantium.
XX
FN WO200065063-A2.
XX
PD 02-NOV-2000.
XX
PF 21-APR-2000; 2000WO-US10886.
XX
PR 22-APR-1999; 99US-0130725.
XX
PA (UYFL ) UNIV FLORIDA.
XX
PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;
PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Allenman AR;
XX
WPI; 2000-679675/66.
DR N-PSDB; AAC68699.
XX
New polynucleotides useful as DNA vaccines for conferring immunity to
rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
XX
Claim 3; Page 33; 63pp; English.
XX
The present sequence is given in a specification relating to nucleic
acid vaccines containing genes to protect animals or humans against
rickettsial diseases caused by a organisms of Rickettsia sp., Ehrlichia
sp., Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
protective against the rickettsial pathogen. The vaccine comprises the
major antigenic protein 1 (MAP1) gene or the major antigenic protein 2
(MAP2) gene of rickettsial pathogens. The nucleic acid vaccines can be
driven by the human cytomegalovirus (HCMV) enhancer-promoter. Cowdria
ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1 and
3gdorf3 may be used in therapeutic and diagnostic applications. The
polypeptides are useful for detecting antibodies associated with
infection by a rickettsial pathogen whilst the polynucleotides may be
used to detect the presence of rickettsial nucleic acids.
XX
Sequence 287 AA;
Query Match 30.7%; Score 459; DB 21; Length 287;
Best Local Similarity 36.1%; Pred. No. 1.1e-35;
Matches 110; Conservative 65; Mismatches 100; Indels 30; Gaps 10;
```

Qy 1 MNKLKFTIINTVLVCLLS-LPNISSSKAINNNAKKYGLYISGOYKPSVSFNSFVKE 59  
Db 1 MNCKKIP--ITSTLISLVSELPVGSFSDVIQEDSNPAGSVYISAKYMPHTASHFGKMSIKE 58  
Qy 60 TNVITKNLIALKXVDSTETKTDSVGSISNPSNFT-IPYTAVFQDNS-VNFNGTIGYTFA 117  
Db 59 DSKNTQTVFGLKQWDGVKTPSDSS--NTNSTIFTEKDYSPRYENNPFPLGPAIGAGYSM- 115  
Qy 118 EGTVEIEGSEVEFDVKNPGGYTLDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRND 177  
Db 116 NGPRIEFSEVETFDVKNLGNKYNNAHMYCAL--DTAAQNSTNGAGLTTSV---MVKNE 170  
Qy 178 GLSIISVIVNVCYDFSLNLSISPIYICGAGVDAIEFEDVLHIKFAVOSKLGIAYSLPN 237  
Db 171 NLTNISLMLNACYDMLDGIPSPVCAGIGTDLVSVINATNPVKLSYQKLGISYSINSE 230  
Qy 238 ISLFASLYYHKMGQNFKNLVQHVLAELASIPKITS-----AVATLNIGYFGGEIG 288  
Db 231 ASIFIGHFHVRVIGNEFK-----DIATLKIFTKTSISNPGFASATLDVCHFGIEIG 282  
Qy 289 ARLTF 293  
Db 283 GRFVF 287  
RESULT 39  
AAU04192  
ID AAU04192 standard; Protein; 287 AA.  
AC AAU04192;  
XX  
XX  
DT 23-OCT-2001 (first entry)  
DE Major antigenic protein 1 (MAP1).  
XX Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;  
KW infection; heartwater; diagnostic.  
XX  
XX Cowdria ruminantium.  
XX  
XX US6251872-B1.  
XX  
XX 26-JUN-2001.  
XX  
XX 17-OCT-1997; 9TUS-0953326.  
XX  
XX 17-OCT-1996; 9GUS-0733230.  
XX  
XX (UVFL) UNIV FLORIDA.  
XX  
XX Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;  
PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;  
XX  
XX WPI; 2001-424487/45.  
XX  
XX N-PSDB; AAS07575.  
XX  
XX New MAP2 genes and polypeptides useful as vaccines for conferring  
PT immunity to human and animal rickettsial diseases, e.g. heartwater, or  
PT as molecular markers in nucleic acid analysis procedures  
XX  
XX Example 1; Column 11-13; 30pp; English.  
XX  
XX The sequence represents the amino acid sequence of major antigenic  
CC protein 1 (MAP1) isolated from Cowdria ruminantium. The MAP  
CC polynucleotides and polypeptides are useful as vaccines for conferring  
CC immunity to rickettsia infection, including Cowdria ruminantium causing  
CC heartwater. The MAP polynucleotides may be used as molecular markers in  
CC nucleic acid analysis procedures, and to produce the MAP polypeptides,  
CC which may be used to raise antibodies that are reactive with the  
CC polypeptides. The nucleic acids may further be used as probes to identify  
CC complementary sequences within other nucleic acid molecules or genomes,  
CC where such probes can be applied to identify or distinguish infectious  
CC strains of organisms in diagnostic procedures or in rickettsial

CC research where identification of particular organisms or strains is  
CC needed.  
XX  
SQ Sequence 287 AA;  
Query Match 30.7%; Score 459; DB 22; Length 287;  
Best Local Similarity 36.1%; Pred. No. 1.1e-35;  
Matches 110; Conservative 65; Mismatches 100; Indels 30; Gaps 10;  
Qy 1 MNKLKFTIINTVLVCLLS-LPNISSSKAINNNAKKYGLYISGOYKPSVSFNSFVKE 59  
Db 1 MNCKKIP--ITSTLISLVSELPVGSFSDVIQEDSNPAGSVYISAKYMPHTASHFGKMSIKE 58  
Qy 60 TNVITKNLIALKXVDSTETKTDSVGSISNPSNFT-IPYTAVFQDNS-VNFNGTIGYTFA 117  
Db 59 DSKNTQTVFGLKQWDGVKTPSDSS--NTNSTIFTEKDYSPRYENNPFPLGPAIGAGYSM- 115  
Qy 118 EGTVEIEGSEVEFDVKNPGGYTLDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRND 177  
Db 116 NGPRIEFSEVETFDVKNLGNKYNNAHMYCAL--DTAAQNSTNGAGLTTSV---MVKNE 170  
Qy 178 GLSIISVIVNVCYDFSLNLSISPIYICGAGVDAIEFEDVLHIKFAVOSKLGIAYSLPN 237  
Db 171 NLTNISLMLNACYDMLDGIPSPVCAGIGTDLVSVINATNPVKLSYQKLGISYSINSE 230  
Qy 238 ISLFASLYYHKMGQNFKNLVQHVLAELASIPKITS-----AVATLNIGYFGGEIG 288  
Db 231 ASIFIGHFHVRVIGNEFK-----DIATLKIFTKTSISNPGFASATLDVCHFGIEIG 282  
Qy 289 ARLTF 293  
Db 283 GRFVF 287  
RESULT 40  
AAU96111  
ID AAU96111 standard; Protein; 284 AA.  
XX  
XX AC AAU96111;  
XX  
XX 02-JUL-2002 (first entry)  
XX  
XX DE Cowdria ruminantium MAP-1.  
XX  
XX KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;  
XX antibacterial.  
XX  
XX OS Cowdria ruminantium.  
XX  
XX PN WO200222782-A2.  
XX  
XX PD 21-MAR-2002.  
XX  
XX PF 12-SEP-2001; 2001WO-US28759.  
XX  
XX PR 12-SEP-2000; 2000US-0660587.  
XX  
XX PA (RERE-) RES DEV FOUND.  
XX  
XX PI Walker DH, Yu X, McBride JW;  
XX  
XX DR WPI; 2002-351882/38.  
XX  
XX PT New recombinant homologous 28 kilodalton immunodominant protein from  
XX Ehrlichia canis, useful for treating Ehrlichia canis infections  
XX  
XX PS Example 3; Figure 3; 106pp; English.  
XX  
XX CC The invention relates to a recombinant homologous 28 kDa immunodominant  
CC protein, p28, (I), of Ehrlichia canis. (i), a 28-kDa antigen preferably  
CC dispersed in a pharmaceutically acceptable carrier, is useful for  
CC inhibiting E. canis infection in a subject. (i) is useful in the  
CC development of vaccines and serodiagnostics that are particularly

CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118  
CC represent the 28-kDa antigen amino acid sequences of the invention.

SQ Sequence 284 AA;

Query Match 30.5%; Score 456.5; DB 23; Length 284;  
Best Local Similarity 34.9%; Pred. No. 1.9e-35;  
Matches 105; Conservative 58; Mismatches 113; Indels 25; Gaps 8;

Qy 1 MNKKLFTIINTVLVCLLS-LPNISSSKAINNNAKYGYLYISGYKPSVFSNFSVKE 59  
Db 1 MNCKKIF--ITSTLSLVFLPGVFSFDVIOEENPVGVSIVISAKYMTASHFGKMSIKE 58

Qy 60 TNVITKNLIAKVDSDIETKTDSVGSISNPSNFTIPYTAVFQDNS-VNENGITIGYTFAE 118  
Db 59 DSRDTKAVFGLKQWDGKTPSGNTNIFTEKDYSEKY---ENNPLGFAGAVGYSM-N 113

Qy 119 GTRVEIEGSYEEDVKNPGGVTLSDAYRYPALAREMKGNSETPKEKVSNSIFHTVMRNDG 178  
Db 114 GPRIEFVSYETFDVRNPGNGYKNDAMHYCALD-----TASSSTAGATTSMVKXNEN 165

Qy 179 LSIISVIVNVCYDFSLNLSISPIYICGAGVDATIEFFDVLHIKPAYOSKLGIAYSLPNSI 238  
Db 166 LTDISLMNACYDMLDGMPSVPVCAGIGTDLVSVINATNPKLSYQKLGISINPEA 225

Qy 239 SLFASLYYHKVGNQFNKLNVOHV-----AELASIPKITSAVATLNIGYFGGEIGARLT 292  
Db 226 SIFIGGHFHRVIGNEFKDIATSKVFTSSGNASSAVSPGFASAI--LDVCHFGIEIGGRFV 283

Qy 293 F 293  
Db 284 F 284

Search completed: July 8, 2003, 10:17:43  
Job time : 74 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 10:13:35 ; Search time 81 Seconds  
(without alignments)  
745.331 Million cell updates/sec

Title: US-10-062-624-40  
Perfect score: 1496  
Sequence: 1 MNKKLKEITNTVLVCLLSL.....AVATLNTGYGGEIGARLTF 293

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1496	100.0	293	2	Q9f477 ehrlichia c
2	1072	71.7	291	2	Q9ACJ0 ehrlichia c
3	1033	69.1	291	2	Q9L6Y4 ehrlichia c
4	1032.5	69.0	300	2	Q9L6Y2 ehrlichia c
5	1026	68.6	297	2	Q9L6Y2 ehrlichia c
6	1008.5	67.4	296	2	Q9ADV5 ehrlichia c
7	1006.5	67.3	298	2	Q9L6V3 ehrlichia c
8	991	66.2	288	2	Q9L6V1 ehrlichia c
9	923	61.7	299	2	Q9ADV6 ehrlichia c
10	820.5	54.8	294	2	Q9ADV4 ehrlichia c
11	789.5	52.8	279	2	Q9ADV8 ehrlichia c
12	789	52.7	285	2	Q9L6V5 ehrlichia c
13	786.5	52.6	279	2	Q9ACJ2 ehrlichia c
14	779.5	52.1	279	2	Q9L6V9 ehrlichia c
15	768	51.3	283	2	Q9L6V7 ehrlichia c
16	757	50.6	289	2	Q9ADV9 ehrlichia c

17	742	49.6	281	2	Q9ADV7
18	737.5	49.3	275	2	Q9L6Y6
19	678.5	45.4	278	2	Q9ADW1
20	671	44.9	295	2	Q9L6Z0
21	611	40.8	289	2	Q9ADW2
22	519	34.7	272	2	Q9L6Z1
23	498	33.3	283	2	Q52103
24	490.5	32.8	280	2	Q9F476
25	480.5	32.1	282	2	Q9WM41
26	478.5	32.0	290	2	Q46332
27	476.5	31.9	284	2	Q9AF41
28	475.5	31.8	282	2	Q9R443
29	471.5	31.5	290	2	Q46333
30	470.5	31.5	280	2	Q52107
31	470.5	31.5	288	2	Q9ZGJ2
32	470.5	31.5	290	2	Q9R864
33	469	31.4	287	2	Q9R425
34	468.5	31.3	280	2	Q85357
35	468.5	31.3	290	2	Q46330
36	466.5	31.2	290	2	Q46324
37	464	31.0	271	2	Q69197
38	462	30.9	281	2	Q9S6H1
39	461.5	30.8	290	2	Q9AEU3
40	459.5	30.7	277	2	Q9R865
41	459.5	30.7	281	2	Q9S6H0
42	459	30.7	287	2	Q46331
43	456.5	30.5	280	2	Q9R855
44	456.5	30.5	284	2	Q46327
45	455	30.4	287	2	Q46329

ALIGNMENTS

RESULT 1

Q9F477 PRELIMINARY; PRT; 293 AA.

AC Q9F477; 01-MAR-2001 (TRENBLrel. 16, Created)  
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)  
DE P28-1 (Major outer membrane protein P30-5).  
GN P28-1 OR P30-5.  
OS Ehrlichia canis.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Ehrlichia.  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=944.  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAKE;  
RX MEDLINE=9242757; PubMed=10225842;  
RA McBride J.W., Yu, Xj, Walker D.H.;  
RT "Molecular cloning of the gene for a conserved major immunoreactive  
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic  
RT antigen.";  
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAKE;  
RX MEDLINE=20432107; PubMed=10974556;  
RA McBride J.W., Yu X.J., Walker D.H.;  
RT "A conserved, transcriptionally active p28 multigene locus of  
RT Ehrlichia canis.";  
RL Gene 254:245-253(2000).  
[3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OKLAHOMA;  
RX MEDLINE=98371112; PubMed=9705412;  
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;  
RT "Cloning and characterization of multigenes encoding the  
RT immunodominant 30-kilodalton major outer membrane proteins of  
RT Ehrlichia canis and application of the recombinant protein for  
RT serodiagnosis.";

Ref don't show the p28

11 V8d vgl

AA-N.A.



```
AC Q9L6Y2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P28-12 (Major outer membrane protein OMP-12).
GN P28-12 OR OMP-12.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RA Yu X.J., McBride J.W., Zhang X., Walker D.H.;
RT "Characterization of the complete transcriptionally active Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Gene 248:29-68(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF230642; AAF73421.1; -.
DR EMBL; U72291; AAK28672.1; -.
DR InterPro; IPR002566; Surface Ag_map4.
DR Pfam; PF01617; Surface Ag_2; 1.
SQ SEQUENCE 300 AA; 33044 MW; 47E6A5A304C000B3 CRC64;

Query Match 69.0%; Score 1032.5; DB 2; Length 300;
Best Local Similarity 66.7%; Pred. No. 1.9e-61;
Matches 200; Conservative 35; Mismatches 58; Indels 7; Gaps 2;

Qy 1 MNKKLFTIINTVLVCLLSLNISSSKAINNNAKYGLYISGQKPSVSVFNSFKVET 60
Db 1 MKKKNQFTITILVCLLSLNASLNTNSTKKQFLGYSGQYKPSVSVFNSFKVET 60

Qy 61 NVITKNLAKKVDVSIETKTDSVIGSNPSNFTIPYTAVFQDQSVNPNFTGYTFAEGT 120
Db 61 NPTKYLAAKKDINSVDFSDVSTAGISYPLNFTPIYTAVFQDQSVNPNFTGYTFAEGT 120

Qy 121 RVEIEGSYEEDVKVNGPGYT-LSDAYRYFALAREMKNSTFKPKVSN-----SIFHTV 173
Db 121 RIEIEGSYEEDVKDPRGYTEIQDAYRYFALARDIDSIPTSPKNTSDHGNSSYKVYHTV 180

Qy 174 MRNGLSIISVIVNVCYDFSLNLSISPIYCGGAGVDIAEFVDVHLHKFAYQSKLGIAYS 233
Db 181 MKNEGLSIISIMVNGCYDFSSDNLISILPYCGGIGVNAIEFFDALLHVKFAQCQKLGITYP 240

Qy 234 LPSNLSLPSIYHYHKVMGNQPKNLNVQHVLAELASTPKITSATVATLNGYFGGEIGARLTF 293
Db 241 LSSNLSLPSAGGYHQVMGNQPKNLNVQHVLAELNAPKVTSAVATLDIGYFGGEIGARLTF 300

RESULT 5
O52102 PRELIMINARY; PRT; 297 AA.
AC O52102;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Major outer membrane protein OMP-1A.
GN OMP-1A.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP PARTIAL SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
```

```
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
RT are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RA Ohashi N., Rikihisa Y., Unver A.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U72291; AAC02935.2; -.
DR InterPro; IPR002566; Surface Ag_map4.
DR Pfam; PF01617; Surface Ag_2; 1.
SQ SEQUENCE 297 AA; 32562 MW; 2ACA25B646EEE141 CRC64;

Query Match 68.6%; Score 1026; DB 2; Length 297;
Best Local Similarity 66.0%; Pred. No. 5.1e-61;
Matches 194; Conservative 40; Mismatches 58; Indels 2; Gaps 2;

Qy 1 MNKKLFTIINTVLVCLLSLNISSSKAINNNAKYGLYISGQKPSVSVFNSFKVET 60
Db 5 MNKKNKFTTISTAMVCLLLPGISFSETINNSAKKQGLYISGQYKPSVSVFNSFKVET 64

Qy 61 NVITKNLAKKVDVSIETKTDSVIGSNPSNFTIPYTAVFQDQSVNPNFTGYTFAEGT 120
Db 65 NVPTKQLIAKKDINSVAVGNSNATGTSNPGNFTPIYTAEFQDQSVNPNFTGYTFAEGT 124

Qy 121 RVEIEGSYEEDVKVNGPGYT-LSDAYRYFALAREMKNSTFKPKVSNISFHTVNRDGL 179
Db 125 RIEIEGFHEKFDVKVNGPGYTVQVQDAYRYFALARDKDGFEKPAE-DTGVYHTVMKNDGL 183

Qy 180 SIISVIVNVCYDFSLNLSISPIYCGGAGVDIAEFVDVHLHKFAYQSKLGIAYS 239
Db 184 SILTMVNVVCYDFSDVDELPLPYICAGMGINAIEFFDALLHVKFAQYQKLGISYQLFTKVN 243

Qy 240 LFASLYYHKVMGNQPKNLNVQHVLAELASTPKITSATVATLNGYFGGEIGARLTF 293
Db 244 LFLDGYHQVMGNQPKNLNVQHVYTLKESPKVTSATVATLDIAYEGGEIGRPTF 297

RESULT 6
Q9ADV5 PRELIMINARY; PRT; 296 AA.
AC Q9ADV5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Major outer membrane protein P30-7.
GN P30-7.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
```

RP SEQUENCE FROM N.A.  
 RC STRAIN=OKLAHOMA;  
 RX MEDLINE=21153566; PubMed=11254561;  
 RA Ohashi N., Rikihisa Y., Unver A.;  
 RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer  
 RT Membrane Protein Multigene Family in Ehrlichia canis and E.  
 RT chaffeensis."; 69:2083-2091(2001).  
 RL Infect. Immun. 69:2083-2091(2001).  
 DR EMBL; AF078553; AAK28693.1; --  
 DR InterPro; IPR002566; Surface\_Ag\_msp4.  
 DR Pfam; PF01617; Surface\_Ag\_2; 1.  
 SQ SEQUENCE 296 AA; 33261 MW; D001CAE71F748E53 CRC64;

Query Match 67.4%; Score 1008.5; DB 2; Length 296;  
 Best Local Similarity 67.8%; Pred. No. 7.5e-60;  
 Matches 200; Conservative 36; Mismatches 54; Indels 5; Gaps 4;

QY 1 MNKLKFTIINTVLVCLLSLPNISSKAINNNKAYGLYISGQYKPSVVSFNSVKET 60  
 DB 5 MNKSQFLIRFLTCMLSLPNSLSK-VNN--EKHSGLYISGQYKPSVVSFNSVKET 61  
 QY 61 NVITKNLALKKQVDSIETKDA-SVGI-SNPSNFTPIPTAVFQDNSVNFNGTGYTFAEG 119  
 DB 62 NFPTKHLIALKQVDSVEIDTSGNTAGISNPSNFTPIPTAEFQDNHTNCGSIGYAFAG 121  
 QY 120 TRVEIEGSYEEDVKNPQGY-TLSDAYRYFALAREMKGNSFTPKKVSNSIFHTWRNDG 178  
 DB 122 PRIETIEYKEDVKNPQGYTVKDAYRYFALAREINISLSLPKQKSGSIYHVVMKNDG 181  
 QY 179 LSIISVINVCVDFSLNLSIPYICGGAGVDAIEFFDVLHFKFAYQSKLGIAYSLSNI 238  
 DB 182 LSILSNVINVCVDFSLNLSIPYICGGAGVDAIEFFDVLHFKFAYQSKLGIAYSLSNI 241  
 QY 239 SLFASLYYHKVMGNQFNLVQHVLAELASIPKITSAVATLNIGYFGGEIGARLTF 293  
 DB 242 NLFIDVYYQVTSNFKNLKVQHVHKLKDPKVTSAVATLDTAYFGSEAGRIIF 296

RESULT 7

Q9L6Y3 PRELIMINARY; PRT; 298 AA.  
 AC Q9L6Y3;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE P28-11 (Major outer membrane protein OMP-1H).  
 GN P28-11 OR OMP-1H.  
 OS Ehrlichia chaffeensis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Anaplasmataceae; Ehrlichia.  
 OX NCBI\_TaxID=945;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ARKANSAS;  
 RA Yu X.J., McBride J.W., Zhang X., Walker D.H.;  
 RT "Characterization of the complete transcriptionally active Ehrlichia  
 RT chaffeensis 28 kDa outer membrane protein multigene family.";  
 RL Gene 248:29-68(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ARKANSAS;  
 RX MEDLINE=21153566; PubMed=11254561;  
 RA Ohashi N., Rikihisa Y., Unver A.;  
 RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer  
 RT Membrane Protein Multigene Family in Ehrlichia canis and E.  
 RT chaffeensis."; 69:2083-2091(2001).  
 RL Infect. Immun. 69:2083-2091(2001).  
 DR EMBL; AF230642; AAF73420.1; --  
 DR EMBL; U72251; AAK28671.1; --  
 DR InterPro; IPR002566; Surface\_Ag\_msp4.  
 DR Pfam; PF01617; Surface\_Ag\_2; 1.  
 SQ SEQUENCE 298 AA; 33026 MW; C2693B131D8CD2D5 CRC64;

Query Match 67.3%; Score 1006.5; DB 2; Length 298;  
 Best Local Similarity 65.2%; Pred. No. 1e-59;  
 Matches 193; Conservative 40; Mismatches 56; Indels 7; Gaps 3;  
 QY 1 MNKLKFTIINTVLVCLLSLPNISSKAINNNKAYGLYISGQYKPSVVSFNSVKET 60  
 DB 1 MNKSMLFTTCTALISLLSPNVVSFGIINNANN-LGIYISGQYKPSVVSFNSVKET 59  
 QY 61 NVITKNLALKKQVDSIETKDA-SVGI-SNPSNFTPIPTAVFQDNSVNFNGTGYTFAEGT 120  
 DB 60 NFFTQQLVALKQDSDVDISTNADSGINPNQFTPIPTIPKFDQNAASFGSALGFFYARGL 119  
 QY 121 RVEIEGSYEEDVKNPQGYT-LSDAYRYFALAREMKGNSFTPKKVS-----NSIEHTVM 174  
 DB 120 RLEMEGSYEEDVKNPQGYTKVDAYRYFALAREMGSQGTCPKHETSGIOPHGIHYTM 179  
 QY 175 RNDGLSIISVINVCVDFSLNLSIPYICGGAGVDAIEFFDVLHFKFAYQSKLGIAYSLS 234  
 DB 180 RNDGVSISSVIINGCYNFTLSNLPISPMVCMGMDALQFDSLHFKFAHQSKLGITVPL 239  
 QY 235 PSNLSLFSALYYHKVMGNQFNLVQHVLAELASIPKITSAVATLNIGYFGGEIGAR 290  
 DB 240 SSNVHLFADSYTHKVGINKFKNLRVQHVYELQVQKVTSAVATLDTIGYFGGEVGR 295

RESULT 8

Q9L6Y1 PRELIMINARY; PRT; 288 AA.  
 AC Q9L6Y1;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE P28-13 (Fragment).  
 GN P28-13.  
 OS Ehrlichia chaffeensis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Anaplasmataceae; Ehrlichia.  
 OX NCBI\_TaxID=945;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ARKANSAS;  
 RA Yu X.J., McBride J.W., Zhang X., Walker D.H.;  
 RT "Characterization of the complete transcriptionally active Ehrlichia  
 RT chaffeensis 28 kDa outer membrane protein multigene family.";  
 RL Gene 248:29-68(2000).  
 DR EMBL; AF230642; AAF73422.1; --  
 DR InterPro; IPR002566; Surface\_Ag\_msp4.  
 DR Pfam; PF01617; Surface\_Ag\_2; 1.  
 FT NON TER 288  
 SQ SEQUENCE 288 AA; 31955 MW; 1ED203FBAED9CAEF CRC64;  
 Query Match 66.2%; Score 991; DB 2; Length 288;  
 Best Local Similarity 66.0%; Pred. No. 1.1e-58;  
 Matches 188; Conservative 39; Mismatches 56; Indels 2; Gaps 2;

QY 1 MNKLKFTIINTVLVCLLSLPNISSKAINNNKAYGLYISGQYKPSVVSFNSVKET 60  
 DB 5 MNKNKFTTISTAMVCLLLPGLISFSETINNSAKKQGLYISGQYKPSVVSFNSVKET 64  
 QY 61 NVITKNLALKKQVDSIETKDA-SVGI-SNPSNFTPIPTAVFQDNSVNFNGTGYTFAEGT 120  
 DB 65 NVPTKQLIALKQDINSVAGSNATGINSNPNFTPIPTAEFQDNVANPENGAVYSFPDSL 124  
 QY 121 RVEIEGSYEEDVKNPQGYT-LSDAYRYFALAREMKGNSFTPKKVSNSIFHTWRNDGL 179  
 DB 125 RIEIEGPEKEDVKNPQGYTQVQDAYRYFALARDLKGDFEPEKAE-DTGVTHTVMKNDGL 183  
 QY 180 SIIISVINVCVDFSLNLSIPYICGGAGVDAIEFFDVLHFKFAYQSKLGIAYSLSNIS 239  
 DB 184 SILSTMVNVCVDFSLNLSIPYICGGAGVDAIEFFDVLHFKFAYQSKLGIAYSLSNIS 243  
 QY 240 LFLASLYYHKVMGNQFNLVQHVLAELASIPKITSAVATLNIGYFG 284



```
Db 244 LFLDGYHQVIGNQFKNLNVNHNVTYTKESPVKVTSVATLDTAYFG 288
RESULT 9
Q9ADV6 PRELIMINARY; PRT; 299 AA.
AC Q9ADV6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Major outer membrane protein P30-8.
GN P30-8.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091 (2001).
DR EMBL; AF078553; AAK28692.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
DR PROSITE; PS00221; MIP; UNKNOWN 1.
DR PROSITE; PS00221; MIP; UNKNOWN 1.
SQ SEQUENCE 299 AA; 33085 MW; 14A29F3A1A3CC42D CRC64;

Query Match 61.7%; Score 923; DB 2; Length 299;
Best Local Similarity 60.9%; Pred. No. 3.8e-54;
Matches 182; Conservative 41; Mismatches 68; Indels 8; Gaps 4;

Qy 1 MNKLKFTIINTVLVCLLSLPNLSKAIANNNAKYYGLYISGQYKPSVSVFNSFKVET 60
Db 1 MNSKSKPTICTSLICLLSSPNTSLSNFIGNST-KHSGLYVSGQYKPSVSVFNSFKVET 59
Qy 61 NVITKNLIALKQVDVSIETK-TDASVGISNPSNFTIPTAVFQNSVNFNGTIGYTFABG 119
Db 60 NHTVQLVALKKDVNSISMNISNGATGISKATNFNLPLYVAFQDNVAFNFGAIGYSLFEQ 119
Qy 120 TRVEIGSYEFDFVKNPGGYTLSDAYRYFALAREM-----KGNSTPKKYS-NSIFHTV 173
Db 120 LNVIEGSEYEFDFVKNPGGYTLSDAYRYFALAREM-----KGNSTPKKYS-NSIFHTV 179
Qy 174 MRNDGLSIISVIVNVCYDFSLNLSISPYICGGAGVDAIEFFDLVHLHKFAYQSKLGAYS 233
Db 180 MRNGLSILSIMINGCYNPLNDLSISPYFCTGIGVDAIEFFDLHLKALQSKIGATYQ 239
Qy 234 LPSNISLFSYIYHKVMGNQFNLVQHVLAELASIPKITSVATLNIIGYFGGEGARLT 292
Db 240 LSDNISLFTNGYHQVIGDQFKNLKVQYIGELKENPKITSVATLNVGFGGEGVRLT 298

RESULT 10
Q9ADV4 PRELIMINARY; PRT; 294 AA.
AC Q9ADV4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Major outer membrane protein P30-6.
GN P30-6.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091 (2001).
DR EMBL; AF078553; AAK28692.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
DR PROSITE; PS00221; MIP; UNKNOWN 1.
DR PROSITE; PS00221; MIP; UNKNOWN 1.
SQ SEQUENCE 294 AA; 32869 MW; 570FAC76A5820867 CRC64;

Query Match 54.8%; Score 820.5; DB 2; Length 294;
Best Local Similarity 56.5%; Pred. No. 2.6e-47;
Matches 166; Conservative 39; Mismatches 84; Indels 5; Gaps 3;

Qy 1 MNKLKFTIINTVLVCLLSLPNLSKAIANNNAKYYGLYISGQYKPSVSVFNSFKVET 60
Db 5 MYKKYKLTAGVLFHMLFLPHVFSFAKNTSNK---LGLYISGQYKPSVSVFNSFKVET 61
Qy 61 NVITKNLIALKQVDVSIETK-TDASVGISNPSNFTIPTAVFQNSVNFNGTIGYTFABG 120
Db 62 NHTVQLVALKKDVNSISMNISNGATGISKATNFNLPLYVAFQDNVAFNFGAIGYSL 121
Qy 121 RVEIGSYEFDFVKNPGGYT-LSDAYRYFALAREMKNSTPKKVSNSIFHTVNRNDGL 179
Db 122 RIEMGSEYEFDAKDLGEYTKIKDAYRYFALVREMHVSLIYPKDN-NTGTHYTVNRNDGI 180
Qy 180 SIISVIVNVCYDFSLNLSISPYICGGAGVDAIEFFDLVHLHKFAYQSKLGATYSLPNSIS 239
Db 181 SISSATNGCYDFPFPSLSLSPYMCIGIGIDAIEFLNALHIKFCACQKGLGYTVSVSPNVN 240
Qy 240 LFASLYYHKVMGNQFNLVQHVLAELASIPKITSVATLNIIGYFGGEGARLT 293
Db 241 LFADGYTHKVMGNKFNLPVQYVNTLEYPRTVSAITLDTLGYLGGEIGIRPIP 294

RESULT 11
Q9ADV8 PRELIMINARY; PRT; 279 AA.
AC Q9ADV8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Major outer membrane protein P30-11.
GN P30-11.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091 (2001).
DR EMBL; AF078553; AAK28692.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
DR PROSITE; PS00221; MIP; UNKNOWN 1.
DR PROSITE; PS00221; MIP; UNKNOWN 1.
SQ SEQUENCE 294 AA; 32869 MW; 570FAC76A5820867 CRC64;

Query Match 54.8%; Score 820.5; DB 2; Length 294;
Best Local Similarity 56.5%; Pred. No. 2.6e-47;
Matches 166; Conservative 39; Mismatches 84; Indels 5; Gaps 3;

Qy 1 MNKLKFTIINTVLVCLLSLPNLSKAIANNNAKYYGLYISGQYKPSVSVFNSFKVET 60
Db 5 MYKKYKLTAGVLFHMLFLPHVFSFAKNTSNK---LGLYISGQYKPSVSVFNSFKVET 61
Qy 61 NVITKNLIALKQVDVSIETK-TDASVGISNPSNFTIPTAVFQNSVNFNGTIGYTFABG 120
Db 62 NHTVQLVALKKDVNSISMNISNGATGISKATNFNLPLYVAFQDNVAFNFGAIGYSL 121
Qy 121 RVEIGSYEFDFVKNPGGYT-LSDAYRYFALAREMKNSTPKKVSNSIFHTVNRNDGL 179
Db 122 RIEMGSEYEFDAKDLGEYTKIKDAYRYFALVREMHVSLIYPKDN-NTGTHYTVNRNDGI 180
Qy 180 SIISVIVNVCYDFSLNLSISPYICGGAGVDAIEFFDLVHLHKFAYQSKLGATYSLPNSIS 239
Db 181 SISSATNGCYDFPFPSLSLSPYMCIGIGIDAIEFLNALHIKFCACQKGLGYTVSVSPNVN 240
Qy 240 LFASLYYHKVMGNQFNLVQHVLAELASIPKITSVATLNIIGYFGGEGARLT 293
Db 241 LFADGYTHKVMGNKFNLPVQYVNTLEYPRTVSAITLDTLGYLGGEIGIRPIP 294
```





```

RC STRAIN=OKLAHOMA;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF078553; AAK28689.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 289 AA; 32460 MW; BEABDA87935D424A CRC64;

Query Match 50.6%; Score 757; DB 2; Length 289;
Best Local Similarity 52.0%; Pred. No. 4.4e-43;
Matches 155; Conservative 46; Mismatches 83; Indels 14; Gaps 6;

QY 1 MNKLKFTIINTVLVCL---LSLPISSSKAINNAKYYGLYISGQYKPSVSVFNSFSV 57
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MNSKTFSLGSLILCLACLPIQSFSSSNVYNT-KHTGLYISGLYKPSVSHSFDSFI 59
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 58 KETNVITKNLIAKKVDVSIETKTDAISVGISNPSNFTIPYTAVPQDNSVNFNGTGYTFA 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 KETYNTTEALFGLKQDIISILRNKETT---QYNNFNVPYTKAFQDDFASFSIAVGYIAN 116
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 118 ECTRVIEGSEYEEFDVKNPGGYTLSDAYRYFALAREMKNSF-TPKEKVSNSIFHTVMRN 176
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117 NGPRIEIGSEYEEFDVKNPGNYTTIDAHRYIALAREKTSYLLSSPKENK-----YVIKN 171
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 177 DGLSIISVTVNVVDFSLNNLSISPYICGAGVDAIEFFDLVHLHKFAVQSKLGIAYSPLS 236
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 NGISVISIINGCYDIISLNDKSVSYICTGFGDDFIEFSAFRFAVQKIGISYSLSS 231
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 237 NISLFSALYYHKVMGNQFKNLNVQVHA-ELASIPKITSATVATNLNIGYFGGEIGARLTF 293
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 232 NIILFTDGYHKVINSQFKNLNVHVVNELTTPKVTSAFLNIEYFGGEFGFLKPIF 289
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 17
Q9ADV7 PRELIMINARY; PRT; 281 AA.
AC Q9ADV7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Major outer membrane protein P30-9.
GN P30-9.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RN J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF078553; AAK28691.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 281 AA; 31525 MW; DE94CF01CD24D967 CRC64;

```

```

Query Match 49.6%; Score 742; DB 2; Length 281;
Best Local Similarity 52.2%; Pred. No. 4.3e-42;
Matches 153; Conservative 41; Mismatches 87; Indels 12; Gaps 3;

QY 1 MNKLKFTIINTVLVCLSLPNISSKAINNAKYYGLYISGQYKPSVSVFNSFSVKET 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MNNKRNFFLIGMSLLINLLLPIDASSMEVYHNTKPKRLYISGQYRCVSHFSKFSVKET 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 NVITKNLIAKKVDVSIETKTDAISVGISNPSNFTIPYTAVPQDNSVNFNGTGYTFAEGT 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 HCNTVQLVGLTKDI-----KVTNNSINTNTSFNFPYVAEFDQDAMSFSGAIGCFYSNF 115
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 121 RVEIGSEYEEFDVKNPGGYTLSDAYRYFALAREMKNSF-TPKEKVSNSIFHTVMRNDCLS 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116 RIEVEASEYEEFDVKNPGEST-TDSYRYFALARGDGNNTPSQK-----FTVMRNDGLL 168
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 181 IISVNVVNVVDFSLNNLSISPYICGAGVDAIEFFDLVHLHKFAVQSKLGIAYSPLSNI 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 169 ISSVMINGCYNVILNDQAEFYICAGLGGDFIEFFNGFHVKLAYQKGVIGISYQIFPEVRL 228
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 241 FASLYYHKVMGNQFKNLNVQVHA-ELASIPKITSATVATNLNIGYFGGEIGARLTF 293
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 FIDGYHKVKNKFKNLHVQVHALAALPKVTSATVATNLNIGYFGCEAGVREIF 281
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 18
Q9L6Y6 PRELIMINARY; PRT; 275 AA.
AC Q9L6Y6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P28-8 (Major outer membrane protein OMP-1X).
GN P28-8 OR OMP-1X.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=21153566; PubMed=11254561;
RA Chashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF230642; AAF73417.1; -.
DR EMBL; U72291; AAK28688.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 275 AA; 30476 MW; 21B2A84CA8ACAE62 CRC64;

Query Match 49.3%; Score 737.5; DB 2; Length 275;
Best Local Similarity 55.3%; Pred. No. 8.3e-42;
Matches 161; Conservative 29; Mismatches 84; Indels 17; Gaps 6;

QY 3 NKLFKFTIINTVLVCLSLPNISSKAINNAKYYGLYISGQYKPSVSVFNSFSVKETNV 62
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2 SKCNFTITGATLIHML-LPNISFPETINNTDKLSGLYISGQYKPGISHSKFSVKEIYN 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 63 ITKNLIAKKVDVSIETKTDAISVGISNPSNFTIPYTAVPQDNSVNFNGTGYTFAEGRV 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 DNIQLIGLRH--NAISTST-----LNINTDFNIPYKVTFFONNITSFSGAIGSDPTGARF 113
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```
Qy 123 EIEGSEYEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVKNSIFHTVMRNDGLSII 182
Db 114 ELEGSYEDVDVDFDCLIKTYRYFALARNPSSPT-----SNN--YTVMRNDGVSIIT 166
Qy 183 SVIVNVCYDFSLNNLSISPYICGAGVDIAIEFFDLHIKFAVQSKLGIAVSLPNSISLFA 242
Db 167 SVIFNGCYDIFLKDLEVPYCVGVGGDFIEFFDLHIKFAVQSKLGIAVSLPNSISLFA 242
Qy 243 SLYTHKVMGNQKNNLVQHVLAELASIPKITSATVATLNIGYFGGIGARLTF 293
Db 227 DGYTHKVMGNQKNNLVQHVLAELASIPKITSATVATLNIGYFGGIGARLTF 275

RESULT 19
Q9ADW1 PRELIMINARY; PRT; 278 AA.
ID Q9ADW1
AC Q9ADW1
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Major outer membrane protein P30-13.
GN P30-13.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;
RA "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=21153566; PubMed=11254561;
RA "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT Chaffeensis.";
RL Infect. Immun. 69:2083-2091 (2001).
DR EMBL; AF078553; AAK28687.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 278 AA; 30829 MW; 3D2F10AD6412C64C CRC64;

Query Match 45.4%; Score 678.5; DB 2; Length 278;
Best Local Similarity 48.6%; Pred. No. 7.3e-38;
Matches 143; Conservative 50; Mismatches 78; Indels 23; Gaps 7;

Qy 1 MNKKLFTIINTVLVCLLSLPNISSKAINNN--AKKYGYLYISQYKPSVSVFNSVK 58
Db 1 MNKK-KSLLIGTILLISLSPNIPKAFSVINHSDISSNVGLYFTGQYRPAVSHESGTVR 59
Qy 59 ETNVTIKNLIAL--KKVDISTETKTDASVGSINSPNFTIPYAVFDQNSVNFNGTIGTTP 116
Db 60 ETNATQQLVSLNTKNNKHIIET-----NFSGIYTAKFQDQNAASFSGAIGYSY 109
Qy 117 AEGTRVIEGSEYEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVKNSIFHTVMRN 176
Db 110 PEGLKFEIEISYERGVKSTKYNQSTNAV-IFALARQT--TSSNPSDNK-----YVVKVN 161
Qy 177 DGLSIISIVNVCYDFSLNNLSISPYICGAGVDIAIEFFDLHIKFAVQSKLGIAVSLPS 236
Db 162 SGLSVASVWINGCYDFSPNNTTISPYICVGVGGDFIEFFDLHIKFAVQSKLGIAVSLPS 221
Qy 237 NISLPSASLYHKVMGNQKNNLVQHVLAELASIPKITSATVATLNIGYFGGIGAR 290
Db 222 RNFIFADMYHKVIGNQKNNLVQHVLAELASIPKITSATVATLNIGYFGGIGAR 275
```

## RESULT 20

```
Q9L6Z0 PRELIMINARY; PRT; 295 AA.
ID Q9L6Z0
AC Q9L6Z0
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Major outer membrane protein OMP-1U).
GN P28-5 OR OMP-1U.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=21153566; PubMed=11254561;
RA "Characterization of the complete transcriptionally active Ehrlichia
RT Chaffeensis 28 kDa outer membrane protein multigene family.";
RL Gene 248:29-68 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=21153566; PubMed=11254561;
RA "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT Chaffeensis.";
RL Infect. Immun. 69:2083-2091 (2001).
DR EMBL; AF230642; AAF73413.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 295 AA; 33616 MW; 4F67BD9B99BE58FE CRC64;

Query Match 44.9%; Score 671; DB 2; Length 295;
Best Local Similarity 46.6%; Pred. No. 2.5e-37;
Matches 139; Conservative 50; Mismatches 93; Indels 16; Gaps 4;

Qy 6 KFTIINTVLVCLLSLPNISS--SKAINNNAKKYGYLYISQYKPSVSVFNSVKETNVI 63
Db 4 KFNFNVNILTLFLPLKPLKFTYANNNTITQKVLGYISQYKPSIPHFKNFSVEENDKV 63
Qy 64 TKNLIALKKVDISTETKTDASVGSINSPNFTIPYAVFDQNSVNFNGTIGYTPAEGTRVE 123
Db 64 V-DLIGLTTDVTYTEHI-----LRDNTKENTHYIAKFNKFNFINFSSAIGYISQGPRL 117
Qy 124 IEGSEYEDVKNPGGYTLSDAYRYFALAREMKGNSFTPK-----EKVSNISIFHTVMR 175
Db 118 IESSYGDVNVNKNYAVQDVNRYFALVREKNGSNFSPKPHETSQPSDSNPKPSFTLMK 177
Qy 176 NDGLSIISIVNVCYDFSLNNLSISPYICGAGVDIAIEFFDLHIKFAVQSKLGIAVSLP 235
Db 178 NNGVFASVWINGCYDFSPNNTTISPYICVGVGGDFIEFFDLHIKFAVQSKLGIAVSLP 237
Qy 236 NISLPSASLYHKVMGNQKNNLVQHVLAELASIPKITSATVATLNIGYFGGIGARLTF 293
Db 238 PSITIFADMYHKVIGNQKNNLVQHVLAELASIPKITSATVATLNIGYFGGIGARLTF 295

RESULT 21
Q9ADW2 PRELIMINARY; PRT; 289 AA.
ID Q9ADW2
AC Q9ADW2
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Major outer membrane protein P30-14.
GN P30-14.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
```

```
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigene encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF078553; AAK28686.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 289 AA; 32889 MW; A8F8647E9A97390C CRC64;

Query Match 40.8%; Score 611; DB 2; Length 289;
Best Local Similarity 43.6%; Pred. No. 2.4e-33;
Matches 129; Conservative 56; Mismatches 101; Indels 10; Gaps 5;

QY 1 MNKLKFTIINTVLVCLSL--PNISSKAINNN--AKYGLYISQYKPSVSFNSVK 58
DB 1 MLQRLNF--INTILAFLLPFPQSTLYIHDEHTQNVGLYISSQYKPSIPYFKNLIE 58
QY 59 ETNVTITKNIALKKDVSITETKTDASVGSINSPNFTIPYTAVFQDQNSVNFNGTIGYTA 118
DB 59 ENSHKTVELMGLANDVTHV-----TEYVLKDKTKNTFPYSAKFRSLNLSAIGYISGQ 113
QY 119 GTRVEIGSYEFEDVKNPGGYTLSDAYRYFALAREMK--GNSPTPKKSVNSIFHTVMRND 177
DB 114 GPRLEIGSYENFDVASCNCPVKANRYIALVRDKKPGNIYPOQSHSHNSMYSYTFIKNN 173
QY 178 GLSILSVIVNVCYDFSLNLSISPYICGGAGVDAIEFFDVLHIKAYQSKGLIAYSLSPN 237
DB 174 GLSILSVINGCYDIAFSNVKLSIPVCAGIGGDFITLFTMHIKAYQSKGLIAYSLSFS 233
QY 238 ISLFASLYHKVMGNQFNKNLVQHVLAELASIPKITSAVATLNTIGYFGGIGARLTF 293
DB 234 ISIFANGHVHKVMDNVFKNLHVKIYIKLQDAPITTSARAKLAIYFGSEVGVRFPV 289

RESULT 22
Q9L621 PRELIMINARY; PRT; 272 AA.
ID Q9L621
AC Q9L621;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P28-4 (Major outer membrane protein OMP-IT).
OS P28-4 OR OMP-IT.
GN Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RA Yu X.J., McBride J.W., Zhang X., Walker D.H.;
RT "Characterization of the complete transcriptionally active Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Gene 248:29-68(2000).
RN [2]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=ARKANSAS;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF230642; AAF73412.1; -
DR EMBL; U72291; AAK28663.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 272 AA; 31234 MW; 40A6A805368B9D2F CRC64;

Query Match 34.7%; Score 519; DB 2; Length 272;
Best Local Similarity 39.7%; Pred. No. 3.1e-27;
Matches 118; Conservative 50; Mismatches 97; Indels 32; Gaps 7;

QY 1 MNKLKFTIINTVLVCL--LSLPNISSKAINNNAK--KYGLYISQYKPSVSFNSFS 56
DB 3 MYNKKHYCYIVTYVITLFFLLPIESLSALIGNVEKDKVSTVSSQYKPSIFHFRNFS 62
QY 57 VKETNVTITKNIALKKDVSITETKTDASVGSINSPNFTIPYTAVFQDQNSVNFNGTIGYTF 116
DB 63 IOESHP-----KXSSEEPK-KIKANLNILKSN---AYNLQFQDNTTSFGTIGY-F 109
QY 117 AEGTRVEIGSYEFEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKSVNSIFHTVMRN 176
DB 110 SKGRLEAEGCYQBFNVKNNSNLSIISNKYHSRIHDE-----YAITN 154
QY 177 DGLSISVIVNVCYDFSLNLSISPYICGGAGVDAIEFFDVLHIKAYQSKGLIAYSLS 236
DB 155 NKLISASIMWNTCYDINSNTSIVPYLCTGIGEDLVGLFTIHFPLAYQGVKSYLNN 214
QY 237 NISLFASLYHKVMGNQFNKNLVQHVLAELASIPKITSAVATLNTIGYFGGIGARLTF 293
DB 215 NILPFSDIYHKVMGNRFKNLYMQVADPNISEETIPIAKLDIGYFGSEIGIRFMF 271

RESULT 23
OS2103 PRELIMINARY; PRT; 283 AA.
ID OS2103
AC OS2103;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Major outer membrane protein OMP-1B.
GN OMP-1B.
OC Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RA MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
RT are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; U72291; AAC02936.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 283 AA; 31017 MW; DCBCF652B771C95D CRC64;
```



```
QY 59 ETVNVI-----TKNLIALKQVD-----SIETKTDSVGSNPSNFTPIPTAVFQDMSVNFN 109
Db 62 ETPVYKDSPTKVKFGLKKEGSITKYSDFTRTD--ISPEGQNNFI-----SGFS 108

QY 110 GTIGYTFAGTRVEIEGSEYEEFDVKNPGGYT--LSDAYRYFALAREMKGNSFTPEKVSNS 168
Db 109 CSIGY-INDGPRVEIEAYQKENPKNPANETDTSYKHGYSR---AETWTDKK----- 159

QY 169 IPHTVMRNDGLSIIISVINVCYDFSLNNLSISPIYICGGAGVDAIEFFDLVHLHKFAYQSKL 228
Db 160 --YVVLTVNGVTFSLSMKNACDYDAEGVFPPIYACAGIGADLSIFDDINLKFPAYQCKI 217

QY 229 GIAYSLPNSISLFASLYYHKMGQNFKNLVQHVAEIASIPKITSAVATLNIGYGGIG 288
Db 218 GISYPITPEISAFIGGYHGVGNKYNKVPKLPVTLTDAPOSTSASVTLDAFYGGELG 277

QY 289 ARLTF 293
Db 278 VRFTF 282

RESULT 26
Q46332 PRELIMINARY; PRT; 290 AA.
AC Q46332;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Major antigenic protein 1.
GN MAPI.
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NYATSANGA;
RX MEDLINE=96400830; PubMed=8807206;
RA Reddy G.R., Sulisano C.R., Harrison R.H., Mahan S.M., Burridge M.J.,
RA Barbet A.F.;
RT "Sequence heterogeneity of the major antigenic protein 1 genes from
RT Cowdria ruminantium isolates from different geographical areas.";
RL Clin. Diagn. Lab. Immunol. 3:417-422 (1996).
DR EMBL; U50834; AAC44145.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 290 AA; 31530 MW; E031A2F712EC5DBB CRC64;

Query Match 32.0%; Score 478.5; DB 2; Length 290;
Best Local Similarity 36.5%; Pred. No. 1.7e-24;
Matches 111; Conservative 61; Mismatches 107; Indels 25; Gaps 9;

QY 1 MNKLKFTIINTVLVCLLS-LPNISSSKAINNNAKKYGLYISGOYKPSVSVFNSFSVKE 59
Db 1 MNCKKIP--ITSLTLSVSLFPGVSDVIQDSSPAGSVIISAKYMPYASHFGKMSIKE 58

QY 60 TNVITKNLIALKQVDSTETKTDSVGSNPSNFTI---PYTAVFQDMS-VNFMGTIGY 114
Db 59 DSKNTQTVPGLKQDVGKVPST-----ENSNSTFTTEKDYSPRYENNPFLGAGRIGY 113

QY 115 TPAETRVIEGSEYEEFDVKNPGGYT--LSDAYRYFALAREMKGNSFTPEKVSNSIFHTVM 174
Db 114 SM-NGPRIEFSEYETFDVKNPGGNYKNDAMHYCAL-----DTAQPTSNQAGTASSVM 166

QY 175 --RNDGLSIIISVINVCYDFSLNNLSISPIYICGGAGVDAIEFFDLVHLHKFAYQSKLGIAYS 233
Db 167 VKNENLTDSLMLNACYDMLDGMPSVPYVCGAGIGTDLVSVINATNPKLISYQKLGISYS 226

QY 234 LPSNISLFASLYYHKMGQNFKNLVQHV-----AELASIPKITSAVATLNIGYGGIGEA 289
Db 227 INPEASIFIGGHRVIGNEFKDIATSKIFTSTKASSIPNPGFASATLDVCHFGIEIGG 286
```

```
QY 290 RLTF 293
Db 287 RFVF 290

RESULT 27
Q9AF41 PRELIMINARY; PRT; 284 AA.
AC Q9AF41;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Major antigenic protein 1.
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALL-3;
RA Bensaïd A., Maillard J.-C., Chantal I., Perez J.-M., Martinez D.;
RT "Cowdria ruminantium major antigenic protein 1 (map1) gene variants
RT are not geographically constrained.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF355200; AAK27216.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 284 AA; 30720 MW; B0D3AEB9F9A809C1 CRC64;

Query Match 31.9%; Score 476.5; DB 2; Length 284;
Best Local Similarity 36.5%; Pred. No. 2.2e-24;
Matches 111; Conservative 58; Mismatches 104; Indels 31; Gaps 8;

QY 1 MNKLKFTIINTVLVCLLS-LPNISSSKAINNNAKKYGLYISGOYKPSVSVFNSFSVKE 59
Db 1 MNCKKIP--ITSLTLSVSLFPGVSDVIQDSSPAGSVIISAKYMPYASHFGKMSIKE 58

QY 60 TNVITKNLIALKQVDSTETKTDSVGSNPSNFTIPTAVFQDMS-VNFMGTIGYTF 117
Db 59 DSRDTKAVFGUKDQDVGKVPSTSGNTSIFTKDYFKY-----DNNPFLGAGIGYSM- 112

QY 118 EGTRVEIEGSEYEEFDVKNPGGYT--LSDAYRYFALAREMKGNSFTPEKVSNS---SIPTV 173
Db 113 NGPRIEFSEYETFDVKNPGGNYKNDAMHYCAL-----DVTSNYTAGAITSVM 160

QY 174 MRNDGLSIIISVINVCYDFSLNNLSISPIYICGGAGVDAIEFFDLVHLHKFAYQSKLGIAYS 233
Db 161 VKNGNLTDISLMLNACYDMLDGMPSVPYVCGAGIGTDLVSVINATNPKLISYQKLGISYS 220

QY 234 LPSNISLFASLYYHKMGQNFKNLVQHV-----AELASIPKITSAVATLNIGYGGIGEA 289
Db 221 INSETPIFGGHRVIGNEFKDITTSKVTNSSGTSVSPGPFASATLDVCHFGIEIGG 280

QY 290 RLTF 293
Db 281 RFVF 284

RESULT 28
Q9R443 PRELIMINARY; PRT; 282 AA.
AC Q9R443;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Major antigenic protein 1 like protein.
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ANTIGUA, WELGEVONDEN, LEMCOT3, AND UMBANEIN;
```



```
RX MEDLINE=9216274; PubMed=10198207;
RA Suleona C.R., Mahan S.M., Barbet A.F.;
RT "The map1 Gene of Cowdria ruminantium Is a Member of a Multigene
RL Family Containing Both Conserved and Variable Genes.";
RL Biochem. Biophys. Res. Commun. 257:300-305(1999).
DR EMBL; AF125279; AAD26353.1; -
DR EMBL; AF125274; AAD26343.1; -
DR EMBL; AF125277; AAD26349.1; -
DR EMBL; AF125278; AAD26351.1; -
DR InterPro; IPR002566; Surface Ag_msp4.
DR Pfam; PF01617; Surface Ag 2; 1.
SQ SEQUENCE 282 AA; 31040 MW; 0CB3C655775CBC90 CRC64;

Query Match 31.8%; Score 475.5; DB 2; Length 282;
Best Local Similarity 36.7%; Pred. No. 2.6e-24;
Matches 112; Conservative 52; Mismatches 98; Indels 43; Gaps 11;

Qy 6 KFTIINTVLVCLLS-LP-----NLSKSKANNNAKKYGLYISQYKPSVSVFNSVK 58
Db 4 KKLIVRSALISLMSPLPQSPAEPVSSNNIGNENAKE--GFYISAKYNPSPHPRKFSAE 61

Qy 59 ETNVI-----TKNIALKKDVID-----SIETKTIDASVGISNPSNFTIPYTAVFQDQNSVNFN 109
Db 62 ETPVYKDSPTKVFGLKEGSIKYSDFTFD--ISEGQNF-----SGFS 108

Qy 110 GTIGYTPAETREVEIEGSEYBFDVKNPGGYT-LSDAYRYFALAREMKGNSFTPKKVSNS 168
Db 109 GSIGY-IMDGPRVEIEAAYQKFNKPNANETDTSYKHYGLSR---AEMADKK----- 159

Qy 169 IFHTVMRNDGLSIISVIVNVCYDPSLNNLSIPYICGGAGVDAIEFPDVLHIKFAVQSKL 228
Db 160 --YVLTNNGVPTSLMNAACYDITAEQVPTPIPVACAGIGADLISIFDDINLKFAVQSKI 217

Qy 229 GIAYSLPNSISLFASLYHKVMGNQFNKLNQHVLAELASIPKITSAVATLNIGYFGGIG 288
Db 218 GISPITEISAFIGYHGVGNKYNKVPKLPVLTLDAPQSIASVLTDAIGFGGELG 277

Qy 289 ARLTFF 293
Db 278 VRFTF 282

RESULT 29
Q46333 PRELIMINARY; PRT; 290 AA.
AC Q46333;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JUN-1996 (TrEMBLrel. 01, Last sequence update)
DE Major antigenic protein 1.
GN MAP1.
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UM BANEIN;
RX MEDLINE=96400830; PubMed=8807206;
RA Reddy R.G., Suleona C.R., Harrison R.H., Mahan S.M., Burridge M.J.,
RT Barbet A.F.;
RT "Sequence heterogeneity of the major antigenic protein 1 genes from
RT Cowdria ruminantium isolates from different geographical areas.";
RL Clin. Diagn. Lab. Immunol. 3:417-422(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UMBANEIN;
RX MEDLINE=9216274; PubMed=10198207;
RA Suleona C.R., Mahan S.M., Barbet A.F.;
RT "The map1 Gene of Cowdria ruminantium is a member of a multigene
RT family containing both conserved and variable genes.";
RL Biochem. Biophys. Res. Commun. 257:300-305(1999).
RN EMBL; U50835; AAC44146.1; -

DR EMBL; AF125278; AAD26352.1; -
DR InterPro; IPR002566; Surface Ag_msp4.
DR Pfam; PF01617; Surface Ag 2; 1.
SQ SEQUENCE 290 AA; 31204 MW; 90E41F34CBDEALAB CRC64;

Query Match 31.5%; Score 471.5; DB 2; Length 290;
Best Local Similarity 37.0%; Pred. No. 5e-24;
Matches 111; Conservative 62; Mismatches 110; Indels 17; Gaps 9;

Qy 1 MNCKLKTINTINTVLVCLLS-LPNISSSKAINNNAKKYGLYISQYKPSVSVFNSVKE 59
Db 1 MNCKKIP--ITSTLISLVSLPGVSFSDVIOEDSNPAGSVYISAKYMTASHFGKMSIKE 58

Qy 60 TNVTIKNLIALKKDVIDSIETKTIDASVGISNPSNFT-IPYTAVFQDQNS-VNFGNTIGYTFA 117
Db 59 DSKNTQTQVFLKDKWDGV--KTPSSDGSNSIIFTEKDYSPKYNPNPLGPAIGYSM- 115

Qy 118 EGTREVEIEGSEYBFDVKNPGGYT-LSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRND 177
Db 116 NGPRIEFESVYETFDVKNPGGNKYNDAMHYCALDTGTPGS--TQATLNSSV--MVKNE 170

Qy 178 GLSIISVIVNVCYDPSLNNLSIPYICGGAGVDAIEFPDVLHIKFAVQSKLGIAYSLPSN 237
Db 171 NLTDIALMLNACYDITIEGMPVSPYVCAGITDVLVSINATNPKLSYQCKLGISVINPE 230

Qy 238 ISLFASLYYHKVMGNQFNKLNQHV----AELASIPKITSAVATLNIGYFGGEIGARLTF 293
Db 231 ASIFIGHFHRVIGNEFKDITTSKIPTSTGKLAATAASPGPASATLDVCHFGIEIGRPFV 290

RESULT 30
O52107 PRELIMINARY; PRT; 280 AA.
ID O52107;
AC O52107;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Major outer membrane protein OMP-1F.
GN OMP-1F.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
RT are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; U72291; AAC02940.1; -
DR InterPro; IPR002566; Surface Ag_msp4.
DR Pfam; PF01617; Surface Ag 2; 1.
SQ SEQUENCE 280 AA; 30731 MW; CCAA6C34E2AF393E CRC64;

Query Match 31.5%; Score 470.5; DB 2; Length 280;
Best Local Similarity 37.5%; Pred. No. 5.6e-24;
Matches 115; Conservative 57; Mismatches 94; Indels 41; Gaps 14;

Qy 1 MNCKLKTINTINTVLVCLLS-LPNISSSKAINNNAKKYGLYISQYKPSVSVFNSVKE 59
Db 1 MNCK-KF-FITTLVSLMSFLPGISFSDAVQND--NVGNGFYISGKYPSVSHFGVFSKQ 57
```

```

QY 60 TNVITKNLIALKKVDSTETKTDASVGI:SNPSN-FTIP-YTAVFQDNS-VNFMGTIGYTF 116
Db 58 ERNITTTGVFLKQDWDG-----STISKNSPNTFNVPNYSFKYENNFPLGFAGAVGY-L 110
QY 117 AEGTRVEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNS-----IPHT 172
Db 111 MNGPRIELEMSEYETEDVKNGQNNYKNDHAKYVALTHNSGG-----KLSNAGDKFVF-- 161
QY 173 VMNDGLSIIISVNVVVCYDFSLNNLSISPIYICGGAGVDAIEFFDVLHIKFAVQSKLGIA 232
Db 162 -LKNEGLDLSLMLNACYDVISGIPFSPYICAGVGTDLISMFEAINPKISYQKGLGVS 220
QY 233 SLPSNISLFSALYYHKMGNOFKNLVQHVLAELASIPKITS-----AVATLINICYFGE 286
Db 221 SISPEASVFGVGHFKHVGNEFRDIP-----AMIPSTLTGNHFTIVLSVCHFGVE 273
QY 287 ICARLTF 293
Db 274 LGGREFN 280

RESULT 31
Q9ZGJ2
ID Q9ZGJ2 PRELIMINARY; PRT; 288 AA.
AC Q9ZGJ2;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 30-kDa major outer membrane protein (p28-8).
GN P30 OR P28-8.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnostics.";
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=JAKE;
RX MEDLINE=99242757; PubMed=10225942;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Immunol. 6:392-399(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=JAKE;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis.";
RL Gene 254:245-252(2000).
DR EMBL; AF078553; AAC68667.1; -.
DR EMBL; AF082744; AAG14362.1; -.
DR InterPro; IPR002566; Surface Ag msp4.
DR Pfam; PF01617; Surface Ag 2; 1.
SQ SEQUENCE 288 AA; 31590 MW; 86DCAECB88E9BFE5 CRC64;

Query Match 31.5%; Score 470.5; DB 2; Length 288;
Best Local Similarity 38.3%; Pred. No. 5.7e-24;
Matches 120; Conservative 51; Mismatches 97; Indels 45; Gaps 12;

QY 1 MNKLKFTINTVLVCLLS-LPNISSSKA-----INNNAKYVGLYISGQKPSVFSFN 54

```

```

Db 1 MNCKRFF--IASALISLMSFLPSVPSIESIHEDNINGN-----FYISAKYMPASHPGV 52
QY 55 FSVKETNVIKNIALKKVDSTETKTDASVGI:SNPSN-FTIP-YTAVFQDNS-VNFMGT 111
Db 53 FSVKEENKTTTGVFLKQDWDGATIKDASSHTIDPSTIFSI:SNYSFKYENNFPLGFAGA 112
QY 112 IGYTFABGTRVEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFH 171
Db 113 IGYSMG-GPRVEFVSEIYFDVKQGNYSKNDHAKYCALSRHTCG---MPOAGHQKXF-- 166
QY 172 TVMNDGLSIIISVNVVVCYDFSLNNLSISPIYICGGAGVDAIEFFDVLHIKFAVQSKLGIA 231
Db 167 VFLKNEGLDLSLMLNACYDITIDSMFPSPYICAGIGSDLVSMFETTPKISYQKGLGVS 226
QY 232 YSLPSNISLFSALYYHKMGNOFKNLVQHVLAELASIPKITSAVA-----TLNI 280
Db 227 YSISPEASVFGVGHFRHVGNEFKD-----IPAITPAGATBIKGTQFTVTTLNI 275
QY 281 GYFGGEIGARLTF 293
Db 276 CHGLELGGREFN 288

RESULT 32
Q93E64
ID Q93E64 PRELIMINARY; PRT; 290 AA.
AC Q93E64;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major antigenic protein MAP1.
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BURKINA FASO;
RX MEDLINE=21539003; PubMed=11682561;
RA Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaid A.,
RA Haydon D.T., van Heerden H., Allsopp B.A.;
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
RT Are Not Geographically Constrained and Show No Evidence of Having
RT Evolved under Positive Selection Pressure.";
RL J. Clin. Microbiol. 39:4200-4203(2001).
DR EMBL; AF368001; AAK98143.1; -.
DR InterPro; IPR002566; Surface Ag msp4.
DR Pfam; PF01617; Surface Ag 2; 1.
SQ SEQUENCE 290 AA; 31560 MW; 94A1A2E9299CC3CA CRC64;

Query Match 31.5%; Score 470.5; DB 2; Length 290;
Best Local Similarity 36.2%; Pred. No. 5.8e-24;
Matches 110; Conservative 60; Mismatches 109; Indels 25; Gaps 9;

QY 1 MNKLKFTINTVLVCLLS-LPNISSSKAINNNAKYVGLYISGQKPSVFSVFSVKE 59
Db 1 MNCKRIF--ITSTLISLVSFLPGVFSVDVIEDQSPAGSVI:SAKYMETASHFGKMSIKE 58
QY 60 TNVITKNLIALKKVDSTETKTDASVGI:SNPSNFTIP-----PYTAVFQDNS-VNFMGTIGY 114
Db 59 DSKNTQTTFVFLKQDWDGKVKVPTS-----ENSNNTSIFTEKDYSPRYENNPLGFAGRIGY 113
QY 115 TFAGTRVEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVM 174
Db 114 SM-NGPRIEFSVYETEDVKNPGGNYKNDHAKYCAL-----ETAQPTSNQCATLASSVM 166
QY 175 -RNDGLSIIISVNVVVCYDFSLNNLSISPIYICGGAGVDAIEFFDVLHIKFAVQSKLGIA 233
Db 167 VKNENLTDLSLMLNACYDILMDGHPVSPYICAGVGTDLVSVINATNPKLISYQKGLGVS 226
QY 234 LPSNISLFSALYYHKMGNOFKNLVQHV-----AELASIPKITSAVATLNIYFGGEIGA 289
Db 227 INPEASIFGHHFRDIGNEFKDIATSKIFTSTSKASSIPNPGFASATLVDVCHFGIEIGG 286

```

```
Qy 290 RLTF 293
Db 287 RFVF 290

RESULT 33
Q9R425 PRELIMINARY; PRT; 287 AA.
ID Q9R425
AC Q9R425;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Major antigenic protein 1.
GN MAP1.
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HIGHWAY, AND CRYSTAL SPRINGS;
RA Suleona C.R., Mahan S.M., Barbet A.F.;
RX MEDLINE=99216274; PubMed=10198207;
RT "The map1 Gene of Cowdria ruminantium Is a Member of a Multigene
  Family Containing Both Conserved and Variable Genes.";
RL Biochem. Biophys. Res. Commun. 257:300-305(1999).
DR EMBL; AF125276; AAD26348.1; -.
DR EMBL; AF125275; AAD26346.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 287 AA; 31104 MW; 15FBE53BC8215F69 CRC64;

Query Match 31.4%; Score 469; DB 2; Length 287;
Best Local Similarity 36.4%; Pred. No. 7.2e-24;
Matches 111; Conservative 65; Mismatches 99; Indels 30; Gaps 10;

Qy 1 MNKLKFTIINTVLVCLLS-LPNTSSSKAINNAKYYGLYISQYKPSVSVFNSVKE 59
Db 1 MNCKKIP--ITSTLISLVSLFPGVFSFSDVIQEDSNPAGSVYISAKYMPYTAHFGKMSIKE 58

Qy 60 TNVITKNLIALKKDVSIEETKTDSVGSISNPSN-IPYTAVFQDNS-VNFNGTIGYTF 117
Db 59 DSKNTQTQVFLGKQDVGKTPSDSS--NTNSTIFTEKDYSPRYENNPFLGAGAGVSM- 115

Qy 118 EGTRVEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRD 177
Db 116 NGPIEFVSVETFDVKNPGGNYKNNAHMYCAL--DTAAQNSTNGAGLTVSV---MWKNE 170

Qy 178 GLSIISVIVNVCYDFSLNNLSISPIYICGAGVDAIEFFDVLHIKPAYOSKLGIAISLPN 237
Db 171 NLTNISLMLNACYDMLDGPVSPVCAGIGTDLVSVINATNPKLSYQKLGISVINS 230

Qy 238 ISLFLSYHKKVMGNQKLVNQHVAELASIPKITS-----AVATNLNIGYFGGEIG 288
Db 231 ASIFIGHGHRVIGNEFK-----DIATLKIFTSTKTSIGNPGFASATLDVCHFGEIG 282

Qy 289 ARLTF 293
Db 283 GRFVF 287

RESULT 34
ID O85357 PRELIMINARY; PRT; 280 AA.
AC O85357;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 28 kDa major surface antigen-4.
GN Ehrlichia chaffeensis.
OS Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.

OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
  Alleman A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
  the tribe Ehrlichiae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
DR EMBL; AF062761; AAC26720.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30743 MW; FBB841DAF08EE4DC CRC64;

Query Match 31.3%; Score 468.5; DB 2; Length 280;
Best Local Similarity 37.5%; Pred. No. 7.6e-24;
Matches 115; Conservative 57; Mismatches 94; Indels 41; Gaps 14;

Qy 1 MNKLKFTIINTVLVCLLS-LPNISSSKAINNAKYYGLYISQYKPSVSVFNSVKE 59
Db 1 MNCK-KF-FITTTLVSLMFLPGISFSDAVQND-NVGGNFYISGKYVPSVSHFGVFSAKQ 57

Qy 60 TNVITKNLIALKKDVSIEETKTDSVGSISNPSN-FTIP-YTAVFQDNS-VNFNGTIGYTF 116
Db 58 ERNTTIGVFLGKQDWDG-----STISKNSPENTFNPNYSFKYENNPFLGAGAGV-L 110

Qy 117 AEGTRVEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNS----IPHT 172
Db 111 MNGPIELENSYETFDVKNQGNVKNDAHKYALTHNSGG-----KLSNAGDKFVF-- 161

Qy 173 VMRNDGLSIISVIVNVCYDFSLNNLSISPIYICGAGVDAIEFFDVLHIKPAYOSKLGIA 232
Db 162 -LKNEGLDLSLMLNACYDVISIEGIPSPYICAGVGTDLISMPAINPKISYQKLGLSY 220

Qy 233 SLPSNISLPLASLYHKVMGNQKLVNQHVAELASIPKITS-----AVATNLNIGYFGGE 286
Db 221 SISPEASVFGVGHFHKVIGNEFRDIP-----AMIEPSTLTGNHFTVITLTVSCHFGVE 273

Qy 287 IGARLTF 293
Db 274 LGGRFNF 280

RESULT 35
ID Q46330 PRELIMINARY; PRT; 290 AA.
AC Q46330;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Major antigenic protein 1.
GN MAP1.
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GARDEL;
RX MEDLINE=96400830; PubMed=8807206;
RA Reddy G.R., Sulsona C.R., Harrison R.H., Mahan S.M., Burridge M.J.,
  Barbet A.F.;
RT "Sequence heterogeneity of the major antigenic protein 1 genes from
  Cowdria ruminantium isolates from different geographical areas.";
RL Clin. Diagn. Lab. Immunol. 3:417-422(1996).
DR EMBL; U50832; AAC44143.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 290 AA; 31204 MW; 17C5E623EE0A1BD CRC64;

Query Match 31.3%; Score 468.5; DB 2; Length 290;
Best Local Similarity 36.7%; Pred. No. 7.9e-24;
```

```
Matches 110; Conservative 63; Mismatches 110; Indels 17; Gaps 9;
QY 1 MNKCLKFTIINTVLVCLLS-LPNISSSKAINNAKYYGLYISGQYKPSVSVFSNFSVKE 59
Db 1 MNCKKIF--ITSTLISLVSFLPGVFSFDVIOEDSSPAGSVISAKYMPNTASHFGKWSIKE 58
QY 60 TNVITKNLIALKQVDSDIETKTDA SVGISNPSNFT-IPYTA VFQDNS-VNFGNTIGYTPA 117
Db 59 DSKNTQTVPGLKNDGCV--KTPSSDSGNNSIIFTEKDYFSKYENPPFLGFGAGAGYSM- 115
QY 118 ECTRVIEIGSYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPEKVSNSIFHTVMRND 177
Db 116 NGPREFEVSYETFDVKNPGYKNDAMHYCALDTGTGPGS--TOGATLNSSV--WVKNE 170
QY 178 GLSIISVIVNVCYDFSLNLSISPYICGGAGVDAIEFFDVLHIKPAYQSKLGIATSLPN 237
Db 171 NLTDIALMLNACYDITLEGMPVSPYVCAGIGTDLVSINATNPKLSYQKGLGISYNPE 230
QY 238 ISLFASLYHKVMGQFNKLNQHV---AELASIPKITSATVATNIGVFGGEIGARLTF 293
Db 231 ASIFIGGHRVIGNEFKDITTSKIFTSTGKLATAASPGFASATLDVCHFGIEIGRVPF 290

RESULT 36
Q46324
ID Q46324 PRELIMINARY; PRT; 290 AA.
AC Q46324;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Major antigenic surface protein.
GN MAP1.
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WEIGEVONDEN STOCK;
RA Fehrsen J., de Villiers E.P., Brayton K.A., van Kleef M.,
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U49843; AA98610.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 290 AA; 31551 MW; 7CECACB0E01670DD CRC64;

Query Match 31.2%; Score 466.5; DB 2; Length 290;
Best Local Similarity 36.4%; Pred. No. 1.le-23;
Matches 110; Conservative 66; Mismatches 105; Indels 21; Gaps 10;
QY 1 MNKCLKFTIINTVLVCLLS-LPNISSSKAINNAKYYGLYISGQYKPSVSVFSNFSVKE 59
Db 1 MNCKKIF--ITSTLISLVSFLPGVFSFDVIOEDSSPAGSVISAKYMPNTASHFGKWSIKE 58
QY 60 TNVITKNLIALKQVDSDIETKTDA SVGISNPSNFT-IPYTA VFQDNS-VNFGNTIGYTPA 117
Db 59 DSKNTQTVPGLKNDGCVKVPSTENTNYS--SLFTEKDYSPRYENNPFLGFGAGAGYSM- 115
QY 118 ECTRVIEIGSYEEFDVKNPGGYTLSDAYRYFAL--AREMKGNSFTPEKVSNSIPTHVMR 175
Db 116 NGPREFEVSYETFDVKNPGYKNDAMHYCALDTAQSGATNGAT----LASSV---MIK 168
QY 176 NDGLSIISVIVNVCYDFSLNLSISPYICGGAGVDAIEFFDVLHIKPAYQSKLGIATSLP 235
Db 169 NENLTNLSMLNACYDITLGDMPVSPYVCAGIGTDLVSINATNPKLSYQKGLGISYN 228
QY 236 SNISLFASLYHKVMGQFNKLNQHV----AELASIPKITSATVATNIGVFGGEIGARL 291
Db 229 SEASIFIGGHRVIGNEFKDITATLKIFTATNKVSTVNPNGPASATLDVCHFGIEIGR 288
QY 292 TF 293
```

```
Db 289 IF 290

RESULT 37
Q69197
ID Q69197 PRELIMINARY; PRT; 271 AA.
AC Q69197;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE P28-20 precursor (Major outer membrane protein P28-1).
DE P28-20 OR P28-1.
GN Ehrlichia chaffeensis.
OS Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RA Yu X.J., McBride J.W., Zhang X., Walker D.H.;
RT "Characterization of the complete transcriptionally active Ehrlichia chaffeensis 28 kDa outer membrane protein multigene family.";
RL Gene 248:29-68(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer Membrane Protein Multigene Family in Ehrlichia canis and E. chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF068234; AAC19134.2; -.
DR EMBL; U72291; AAK28674.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 271 P28-20.
SQ SEQUENCE 271 AA; 30050 MW; 84EDF665DF3DB61C CRC64;

Query Match 31.0%; Score 464; DB 2; Length 271;
Best Local Similarity 38.3%; Pred. No. 1.5e-23;
Matches 115; Conservative 49; Mismatches 98; Indels 38; Gaps 12;
QY 1 MNKCLKFTIINTVLVCLLS-LPNISSSKAINNAKYYGLYISGQYKPSVSVFSNFSVKE 59
Db 1 MNKY-KF-VVGVALATLLSLFDPNSFSDANVPEGRK--GFVGTQYKGVVPNFSNFSAE 56
QY 60 T-NVITKNLIALKQVDSDIETKTDA SVGISNPSNFTIPYTA VFQDNSVNFGNTIGYTFAE 118
Db 57 TLPGLTKSIFALGDKSS-----ISDHAGFTQAYNPYASNFAGFGGVIGY-YVN 105
QY 119 GTRVEIGSYEEFDVKN---PGGYTLSDAYRYFALARE--MKGNSFTPEKVSNSIPTHV 173
Db 106 DFRVEFGAYENFEPFQWYPEG---GESHKFFALSRETVQDNKF-----IV 150
QY 174 MRNDGLSIISVIVNVCYDFSLNLSISPYICGGAGVDAIEFFDVLHIKPAYQSKLGIAT 233
Db 151 LENDGVLDKSLNVNFCYDIAHSGISPLAPYMCAGVADYIKFLGSLPKFSQVAFGVNYP 210
QY 234 LPSNISLFASLYHKVMGQFNKLNQHV--OHVAELASIPKITSATVATNIGVFGGEIGARL 292
Db 211 VSVNMLFGGYYHKVIGNRYVERVEIAYHPATLTNVPKTTASATLDTDFGVEGVGRFT 270
```



Db 115 FEVYETEDVKNPGGNYKNDHMYCAL-----DTAQPTSNQCATLASSVMVKENLTDI 168  
Qy 183 SVIVNVCYDFSNNLSISPYICGGAGVDIAIEFFDLVHIKFAYQSKLGIAYSLSNLSLFA 242  
Db 169 SLMLNACYDIMLDGMPVSPYVCAGIGTDLVSVINATNPKLSYQKLGISYSINPEASIFI 228  
Qy 243 SLYYHKVMGNQFNLNVOHV-----AELASIPKITSATLNLNIGYFGGEI 287  
Db 229 GGHFHRVIGNEFKDIATSKIPTSTSKASSIPNPGFASATLDVCHFGIEI 277

Search completed: July 8, 2003, 10:19:41  
Job time : 83 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 10:15:30 ; Search time 41 Seconds  
(without alignments)  
687.009 Million cell updates/sec

Title: US-10-062-624-40  
Perfect score: 1496  
Sequence: 1 MNKCLKTIINTVLVCLLSL.....AVATNLNGYGGEGIGARLTF 293

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 73.\*

1: Pirl.\*

2: Pirl2.\*

3: Pirl3.\*

4: Pirl4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	468.5	31.3	280	2 JE0217	28k surface antige
2	456.5	30.5	284	2 I40882	major antigenic pr
3	441.5	29.5	278	2 JE0216	28k surface antige
4	427	28.5	286	2 JE0219	28k surface antige
5	426	28.5	287	2 JE0220	28k surface antige
6	422	28.2	276	2 JE0218	28k surface antige
7	294	19.7	282	2 I39648	major surface prot
8	173	11.6	133	2 JE0221	28k surface antige
9	110.5	7.4	239	2 AH0541	probable outer mem
10	109	7.3	663	2 A86626	hypothetical prote
11	108.5	7.3	1819	2 D79033	uncharacterized pr
12	108	7.2	1441	2 A86695	prophage pil prote
13	107	7.2	401	2 T17515	hypothetical prote
14	105.5	7.1	1152	2 AC1347	probable peptidogl
15	105	7.0	810	2 G86797	prophage p13 prote
16	103	6.9	3194	2 D71917	toxin-like outer m
17	102	6.8	1227	2 C97033	uncharacterized pr
18	102	6.8	1398	2 T20434	hypothetical prote
19	101	6.8	536	2 A99283	hypothetical prote
20	101	6.8	1413	2 B82877	conserved hypothet
21	100.5	6.7	465	2 E70193	purine-binding che
22	100	6.7	449	2 D87682	OmpA family protei
23	100	6.7	1943	2 B64596	toxin-like outer m
24	99	6.6	650	2 S22835	alpha-agglutinin -
25	99	6.6	1441	2 B86807	hypothetical prote
26	98.5	6.6	492	2 B64642	cell division prot
27	98.5	6.6	928	2 C81255	probable lipoprote
28	98.5	6.6	978	2 D81411	probable lipoprote
29	98.5	6.6	1190	2 A82615	surface protein XF

30	98.5	6.6	2893	2 A64556	toxin-like outer m
31	98	6.6	386	2 S58755	ribosomal protein
32	98	6.6	820	2 T17519	cell surface antig
33	98	6.6	1039	2 T30856	protein P2 - Strep
34	97.5	6.5	722	2 F87662	TonB-dependent rec
35	97	6.5	407	2 H87341	OmpA family protei
36	97	6.5	444	2 S54039	hypothetical prote
37	97	6.5	1151	2 AG1717	probable peptidogl
38	96.5	6.5	249	2 A97860	hypothetical prote
39	96.5	6.5	290	2 A11730	Lactobacillus delb
40	96	6.4	373	2 F81438	probable periplasm
41	96	6.4	380	1 S42734	hypothetical prote
42	96	6.4	750	2 D86245	hypothetical prote
43	96	6.4	5005	2 F82884	hypothetical prote
44	95.5	6.4	334	2 B89812	outer membrane ass
45	95.5	6.4	835	2 E71691	

## ALIGNMENTS

### RESULT 1

JE0217

28k surface antigen 4 - Ehrlichia chaffensis

N/Alternate names: MAPI

C/Species: Ehrlichia chaffensis

C/Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 17-Mar-1999

C/Accession: JE0217

R/Reddy, G.R.; Suleona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrige, M.J.; Allenan, A.R.

Biochem. Biophys. Res. Commun. 247, 636-643, 1998

A/Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe

A/Reference number: JE0216; MUID:98321180; PMID:9647746

A/Accession: JE0217

A/Molecule type: DNA

A/Residues: 1-280 <RED>

A/Cross-references: GB:AF062761

Query Match 31.3%; Score 468.5; DB 2; Length 280;  
Best Local Similarity 37.5%; Pred. No. 3.8e-27;  
Matches 115; Conservative 57; Mismatches 94; Indels 41; Gaps 14;

Qy	1	MNNKLFKIINTVLVCLLS-LPNISSKAINNAKKYGLYISGOYKPSVSFNSFSVKE	59
Db	1	MNCK-KF-FITTTLSLMSLPGISFSDAVQND-NVGGNFYISGKYPSVSHEGVSFAKQ	57
Qy	60	TNVTIKNLIALKKDVDSIETKTASVIGSNPSN-FTIP-YTAVFODNS-VNENGTCGYTF	116
Db	58	ERNTTTGVFGKQDWG-----STISKNSPENTFNVPNSFYKNNPFLGAGAVGY-L	110
Qy	117	AEGTRVEIGSYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKEKVSNS----	172
Db	111	MNGPRIELEMSYETFDVQNGNNYKNDHKYVALTHNSGG-----KLSNAGDKPVP--	161
Qy	173	VMRNDGLSIISIVNVCYDFSLNLSISPICGAGVDAIEFFDLVHLKIFAYSKLGIAIY	232
Db	162	-LKNGLGLDLSLMLNACYDVISGIPFSPICAGVGTDLISMFEAINPKISYQKGLSLY	220
Qy	233	SLPSNTSLRSLYHYKVMGNQFNQVLAELASIPKITS-----AVATNLNGYGGGE	286
Db	221	SISPEASVFGGHFKVIGNEFRDIP-----AMIPSTLTGNGHPTITVLSVCHFGVE	273
Qy	287	IGARLTF 293	
Db	274	LGGRNF 280	

### RESULT 2

I40882

major antigenic protein - heartwater rickettsia

C/Species: Cowdria ruminantium (heartwater rickettsia)

C/Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 08-Oct-1999

C/Accession: I40882; S42827

R/van Vliet, A.H.; Jongejan, F.; van Kleef, M.; van der Zeijst, B.A.

Infect. Immun. 62, 1451-1456, 1994  
A:Title: Molecular cloning, sequence analysis, and expression of the gene encoding the  
A:Reference number: 140882; MUID:94178956; PMID:8132352  
A:Accession: J02016  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-284 <RES>  
A:CROSS-references: EMBL:X74250; NID:9454266; PIDN:CA52309.1; PID:9454267  
C:Genetics:  
A:Gene: map1

Query Match 30.5%; Score 456.5; DB 2; Length 284;  
Best Local Similarity 34.9%; Pred. No. 2.9e-26;  
Matches 105; Conservative 58; Mismatches 113; Indels 25; Gaps 8;  
QY 1 MNKKLFTIINTVLVCLLS-LPNISSKAI-NNNAKKYGLYISGOYKPSVSVFSNFSVK 59  
DB 1 MNCKKIP--ITSTLSVSLFPGVSFSDVIQENNPVGSVYISAKYMPATASHFGKMSIKE 58  
QY 60 TTVITKNIALKKQVDSIETKTDA SVGISNPSNFTPIPTAVFQDNS-VNFNGTIGYTAE 118  
DB 59 DSRDTKAVFGLKQWGVKPTSGNTNSIFTEKDYFKY-----ENNPFGLFAGAVGYSM-N 113  
QY 119 GTRVEIGSYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIPTHVNRNDG 178  
DB 114 GRIIEFVSYEYFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIPTHVNRNDG 165  
QY 179 LSIISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDLHIKPAYOSKLGIAYSLSN 238  
DB 166 LTDLSMLNACYDILMDGMPVSPYVACAGIGTDLVSINATNPKLSYQKGLGISYNPEA 225  
QY 239 SLFASLYHKVMGNQKPNLVQHV-----AELASIPKITSAVATLNIGYFGEIGARLT 292  
DB 226 SIFIGGHRVIGNEFRDIPKTSVTSNGASSAVSPGPASAI--LDVCHFGIEIGGRFV 283  
QY 293 F 293  
DB 284 F 284

RESULT 3  
J02016  
28k surface antigen 3 - Ehrlichia chaffensis  
N:Alternate names: MAP1  
C:Species: Ehrlichia chaffensis  
C:Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 17-Mar-1999-  
C:Accession: J02016  
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.  
Biochem. Biophys. Res. Commun. 247, 636-643, 1998  
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe  
A:Reference number: J02016; MUID:98321180; PMID:9647746  
A:Accession: J02016  
A:Molecule type: DNA  
A:Residues: 1-278 <RED>  
A:CROSS-references: GB:AF062761

Query Match 29.5%; Score 441.5; DB 2; Length 278;  
Best Local Similarity 37.4%; Pred. No. 3.6e-25;  
Matches 111; Conservative 56; Mismatches 107; Indels 23; Gaps 12;  
QY 1 MNKKLFTIINTVLVCLLS-LPNISSKAI-NNNAKKYGLYISGOYKPSVSVFSNFSVK 58  
DB 1 MNCK-KF-FITLALVSLMSFLPGISFSDVPQDGNISGNF--YVSGKYMPSASHFGNFSK 56  
QY 59 ETNVTITKNIALKKQVDSIETKTDA SVGISNPSNFTPIPTAVFQDNS-VNFNGTIGYTAE 117  
DB 57 BEKNPTVALYGLKQWEGISSSHNDHFNKNG-----YSPKYENNPFLGFAAGIYSGM 111  
QY 118 EGRVEIGSYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIPTHVNRND 177  
DB 112 -GPRVEFVSYEYFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIPTHVNRND 163  
QY 178 GLSIIISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDLHIKPAYOSKLGIAYSLSN 237

DB 164 GLDISFPLNACYDIINESIPLSPIYICAGVGTDLISFEATNPKISYQKGLSISINPE 223  
QY 238 ISLFASLYHKVMGNQKPNLVQHVVAELASIPKITSAVATLNIGYFGEIGARLT 293  
DB 224 ASVEIGGHRVIGNEFRDIPKTSVTSNGASSAVSPGPASAI--LDVCHFGIEIGGRFNF 278  
RESULT 4  
J02019  
28k surface antigen 2 - Ehrlichia chaffensis  
N:Alternate names: MAP1  
C:Species: Ehrlichia chaffensis  
C:Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 17-Mar-1999-  
C:Accession: J02019  
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.  
Biochem. Biophys. Res. Commun. 247, 636-643, 1998  
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe  
A:Reference number: J02016; MUID:98321180; PMID:9647746  
A:Accession: J02019  
A:Molecule type: DNA  
A:Residues: 1-286 <RED>  
A:CROSS-references: GB:AF062761  
Query Match 28.5%; Score 427; DB 2; Length 286;  
Best Local Similarity 36.7%; Pred. No. 4.3e-24;  
Matches 108; Conservative 57; Mismatches 109; Indels 20; Gaps 12;  
QY 6 KFTIINTVLVCLLS-LPNISSKAI-NNNAKKYGLYISGOYKPSVSVFSNFSVKETNVI 63  
DB 5 KF-FITLALVSLMSFLPGISLSDVPQDGNISGNF--YISGKYMPSASHFGVFSKAEKENT 61  
QY 64 TKNLIALKKQVD-SIETKTDA SVGISNPSNFTPIPTAVFQDNSV-NFNGTIGYTAE 120  
DB 62 TVGVGIEQWDRVCVIRTTLS-----DIFTVPNYSFKYENNPFLGFAAGIYSGM-DGP 114  
QY 121 RVEIGSYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIPTHVNRNDG 180  
DB 115 RIELEVSYEAFDVKNQGNVKNYKNEAHRYALS-HULGTETQIDGAGSASF---LINEGLL 170  
QY 181 IISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDLHIKPAYOSKLGIAYSLSN 240  
DB 171 DKSFPLNACYDVISGIPFSPYICAGIDLVSMFEAINPKISYQKGLSISINPEASV 230  
QY 241 FASLYHKVMGNQKPNLVQHVVAELASIPKITSAVATLNIGYFGEIGARLT 293  
DB 231 FIGGHRVIGNEFRDIPKTSVTSNGASSAVSPGPASAI--LDVCHFGIEIGGRFNF 284

RESULT 5  
J02020  
28k surface antigen 1 - Ehrlichia canis  
C:Species: Ehrlichia canis  
C:Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 21-Jul-2000  
C:Accession: J02020  
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.  
Biochem. Biophys. Res. Commun. 247, 636-643, 1998  
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe  
A:Reference number: J02016; MUID:98321180; PMID:9647746  
A:Accession: J02020  
A:Molecule type: DNA  
A:Residues: 1-287 <RED>  
A:CROSS-references: GB:AF062762; NID:93327964; PIDN:AAC26721.1; PID:93327965  
Query Match 28.5%; Score 426; DB 2; Length 287;  
Best Local Similarity 34.6%; Pred. No. 5.1e-24;  
Matches 110; Conservative 54; Mismatches 98; Indels 56; Gaps 13;  
QY 1 MNKKLFTIINTVLVCLLS-LPNISSKAI-NNNAKKYGLYISGOYKPSVSVFSNFS 56  
DB 1 MKYKKTFTV--TALLVLTSTFTHFPFYSAPAR-----ASTIHNFYISKYNMTASHFGIFS 53  
QY 57 VKETNVTITKNIALKKQVD-SIETKTDA--SVGISNPSNFTPIPTAVFQDNS-VNFNGT 112



Db 54 AKESQSTKVLVGLDQRLSHNIINNDVAKSLKVN-----YSFYKKNPFIFGFAGAI 106  
Qy 113 GYTFAGTRVIEGSEYEFVKNPGGYTLSDAYRYFALARE-----MKGNS---FTPKKVK 165  
Db 107 GYSIG-NSRIEVLVSHEDTKNFGNNYLNDSHKYCALSHGSHICSDDGSDWTAKTDK 165  
Qy 166 SNSIFHTVMRNDGLSIISVIVNVCYDFSLNLSISPYICGAGVDAIEFFDVLHIKPAYQ 225  
Db 166 -----FVLLKNEGLDVSFMLNACYDITTKMPFSPYICAGIGTDLISMFETQNKISY 220  
Qy 226 SKGLIAYSLPNSISLFASLYHKMGQFK-----NLNVQHVLAELASIPKITSVAV 275  
Db 221 GKLGNTYINSRVSVFAGGHFKVIGNEFGIPILLPDGSKIRVQ-----SAT 269  
Qy 276 ATLNIYFGGIGARLTF 293  
Db 270 VTLDVCHFLGIGSRFF 287

## RESULT 6

28k surface antigen 5 - Ehrlichia chaffensis  
N:Alternate names: Map1  
C:Species: Ehrlichia chaffensis  
C:Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 17-Mar-1999  
C:Accession: JE0218  
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; BurrIDGE, M.J.; Allemen, A.R.  
Biochem. Biophys. Res. Commun. 247, 636-643, 1998  
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe  
A:Reference number: JE0216; MUID:98321180; PMID:9647746  
A:Accession: JE0218  
A:Molecule type: DNA  
A:Residues: 1-276 <RED>  
A:Cross-references: GB:AF062761

Query Match 28.2%; Score 422; DB 2; Length 276;  
Best Local Similarity 35.5%; Pred. No. 9.5e-24;  
Matches 107; Conservative 57; Mismatches 99; Indels 38; Gaps 13;  
Qy 1 MNKLKFTIINTVLVCLL-SLPNLS-----KATNNNAKKYGLYISGOYKPSVVSFNF 55  
Db 1 MNYKKVF--ITSALISLISLPGVSDPDAGSGINGN-----FYISGKMPSASHFGVP 52  
Qy 56 SVKETNVTITKNLIALKQDVDSIETKTDASVGISNPSN-FTIP-YTAVFQDNS-VNFGTI 112  
Db 53 SAKERNTTVGVFLGKQWDG-----SAISNSPNDVFTVSNYSFKYNNPFLGFAGAI 106  
Qy 113 GYTFAGTRVIEGSEYEFVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVKNSIFHT 172  
Db 107 GYSM-DGPRIEVLVSYETFDKQGNKYKNEAHRYCALSH-----NSAADSSASNNF--V 159  
Qy 173 VMRNDGLSIISVIVNVCYDFSLNLSISPYICGAGVDAIEFFDVLHIKPAYQSLGIAY 232  
Db 160 FLKNEGLDVSFMLNACYDVGEGIPSYICAGIGTDLVSMFEATPKISYQKGLSY 219  
Qy 233 SLPSNISLFASLYHKMGQFNLV-----QHVLAELASIPKITSVAVATINIGYFGGEI 287  
Db 220 SISPEASVFIGGHFKVIGNEFRDIPITIPGSLAGKNYP-----AIVLDVCHFGIEM 275  
Qy 288 G 288  
Db 276 G 276

## RESULT 7

major surface protein 4 - Anaplasma marginale  
C:Species: Anaplasma marginale  
C:Date: 19-Jul-1996 #sequence\_revision 19-Jul-1996 #text\_change 08-Oct-1999  
C:Accession: I39648  
R:Oberle, S.M.; Barbet, A.F.  
Gene 136, 291-294, 1993

A:Title: Derivation of the complete map4 gene sequence of Anaplasma marginale without cl.  
A:Reference number: I39648; MUID:94124017; PMID:8294020  
A:Accession: I39648  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-282 <RES>  
A:Cross-references: GB:L01987; NID:G142174; PIDN:AAC36877.1; PID:G142175  
C:Genetics:  
A:Gene: msp4

Query Match 19.7%; Score 294; DB 2; Length 282;  
Best Local Similarity 26.8%; Pred. No. 2.4e-14;  
Matches 82; Conservative 57; Mismatches 129; Indels 38; Gaps 9;

Qy 1 MNKLKFT--IINTVLVCLLSPNLSISKAINNNNAKKYGL-----YISGOYKPSVSFV 52  
Db 1 MNYRELFTGGLSAATVCACSLLVSGAVVASPSHEVASEGVGMSGFVGAAYSPAPSV 60  
Qy 53 SNFSVKETNVTITKNLIALKQDVDSIETKTDASVGISNPSNFT-IPYTAVFQDNSV-NFNG 110  
Db 61 TSFDMRESSKETSIVRGYDKSI-----ATIDVSVPANFSKGYTFAFSKNLITSPDG 112  
Qy 111 TIGYTFAGTRVIEGSEYEFVKNPGGYTLSDAYRYFALARE---MKGNSFTPKKVKSN 167  
Db 113 AVGISLG-GARVELEASRRPATLADQYAKGAESLAATRDANITETNYF----- 163  
Qy 168 SIFHTVMRNDGLSIISVIVNVCYDFSLNLSISPYICGAGVDAIEFFDVLHIKPAYQSK 227  
Db 164 -----VVKIDEINTSVMLNGCYDVLHTDLVPVSPYVAGIGASPVDISKQVTTKLAYRGK 218  
Qy 228 LGTAYSLPNSISLFASLYHKMGQFNLVQHVLAELASIPKITSVAVATINIGYFGGEI 287  
Db 219 VGISYQFTPEISLVAGGFYHGLFDESYKIDIPAHNSVRFSGEAK---ASVKAHIADYGFNL 275  
Qy 288 GARLTF 293  
Db 276 GARPLF 281

## RESULT 8

JE0221  
28k surface antigen 2 - Ehrlichia canis  
C:Species: Ehrlichia canis  
C:Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 21-Jul-2000  
C:Accession: JE0221  
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; BurrIDGE, M.J.; Allemen, A.R.  
Biochem. Biophys. Res. Commun. 247, 636-643, 1998  
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe  
A:Reference number: JE0216; MUID:98321180; PMID:9647746  
A:Accession: JE0221  
A:Molecule type: DNA  
A:Residues: 1-133 <RED>  
A:Cross-references: GB:AF062762; NID:G3327964; PIDN:AAC26722.1; PID:G3327966

Query Match 11.6%; Score 173; DB 2; Length 133;  
Best Local Similarity 38.1%; Pred. No. 7e-06;  
Matches 53; Conservative 18; Mismatches 58; Indels 10; Gaps 6;

Qy 1 MNKLKFTIINTVLVCLLSPNLSISKAINNNNAKKYGLYISGOYKPSVVSFNSFKET 60  
Db 1 MNCKKVF--ISALISYIFLPNVSVNPNVYGNM--MYGNFYISGYMPSVPHFGIFSABEE 58  
Qy 61 NVITKNLIALKQDVDSIETKTDASVGISNPSNFTI-PYTAVFQDNS-VNFGNTIGYTFAE 118  
Db 59 KKKTIVYGLKEN-----WAGDAISSQSPDNTIRNYSFKYASKNKFLGFAVAGYSIG- 112  
Qy 119 GTRVEIGSEYEFVKNPG 137  
Db 113 SPRIEVMSEYAEFDVKNQG 131

## RESULT 9

AH0541

Query Match	7.3%	Score 109;	DB 2;	Length 663;
Best Local Similarity	23.5%;	Pred. No. 2.6;		
Matches 74;	Conservative 43;	Mismatches 120;	Indels 78;	Gaps 17;
Qy	1	MNNKLKFTIINTVLCLSLENISSKAINNNAKYTYGLYISGQ-YKPSVSVFNSFYSKE	59	
Dd	1	MNKVIKFLVLLFVISISSANIYKNPNFDVKIQIKENVDVQDFTQYLPS-SISDD-----	54	
Qy	60	TNVTIKNLIALKGVDSTIETKTDAVV-----GISNPSSN-----FTIPYTAVFDQNSVN	107	

Query Match	7.3%;	Score 108.5;	DB 2;	Length 1819;
Best Local Similarity	22.0%;	Pred. No. 10;		
Matches	82;	Conservative 52;	Mismatches 120;	Indels 119; Gaps 20;
Qy	3	NKLKFTINTVLCLSLPNISSKAIN-----NNAKKYGLV-----	40	
Db	890	NSLGITGVTTENTYINL-NIKESSVPNASDIOTKNALISVGYEENKGEATVDVYT	948	
Qy	41	---ISGQYKPSVSFNSFKETNVTITKNLIALKQDVDSIETKTDSVSGISNSPNTIPY	97	
Db	949	SLGITGVTKENI-IFINTVIKEGYF-DLTSLKSSVEVLEEKYEAFVKITSGKAVVG DY	1005	
Qy	98	TA-----VFQDN--SVNFNCTIGTYAECT--RVEI-----EGSYEEFDV	133	
Db	1006	TKVGIKDVTEENIAYINLNIIDLQNCETPTVQARIELVKQYNSYIVISKGSATVEDYEA	1065	
Qy	134	KNPGGYTLDAYRYFALAREMKGNSTPPEKVSNSIFHTWMRNDGLSIISVIVNVCYDFS	193	
Db	1066	INITGVT-KDNIKY--INCELKGTNPAESAEVQAKV-----DEVVKVYEELVRI-----N	1112	
Qy	194	LNNLSISPYIC-CGAGVDAIE-----FFDVLHIK-----FAYQSKLG-----	229	
Db	1113	LGTASVNDYTAIGVSGVADINIALVNQDMKBEKGLSVEEIKQRIADAVFSLOTVLGRINS	1172	
Qy	230	-----IAYSLPNI--SLFASLYYHKVMGNQFKNLNVQHVAAELASI	268	
Db	1173	EGILSDYEAAAGVGVTTNENLAY-VNTNIKGLFESLEDVKASVNEI-----VAQVEIS	1224	
Qy	269	PKITSATVATLNIG	281	
Db	1225	IKIEQIVQRI NLG	1237	

RESULT 12  
 A86685  
 prophage p13 protein 46, tail component [imported] - Lactococcus lactis subsp. lactis (s  
 C;Species: Lactococcus lactis subsp. lactis  
 C;Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C;Accession: A86685  
 R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich  
 Genome Res. 11, 731-753, 2001  
 A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s  
 A;Reference number: A86625; MUID:21235186; PMID:11337471  
 A;Accession: A86685  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-1441 <STO>  
 A;Cross-references: GB:AB005176; PID:g12723361; PIDN:AAK04579.1; GSPDB:GN00146  
 A;Experimental source: strain IL1403  
 C;Genetics:  
 A;Gene: p146

Query Match 7.2%; Score 108; DB 2; Length 1441;  
 Best Local Similarity 25.4%; Pred. No. 8.2;  
 Matches 64; Conservative 43; Mismatches 93; Indels 52; Gaps 15;  
 Qy 14 LVCLLSLPNI---SSSKAINNAKYYGLYISGQYKPSVFSNFSVKETNVITKNIALK 71  
 Db 1214 MINTLSANNIEANSIKADKIDAKNL--TITDGEF---VSTTTNGEPTTSTEIKDNHIAIS 1268  
 Qy 72 KDVDSEITKTDASVGISN--PSNFTIPTAVFQDNSVNFNGTIGYTFAGTRVEIGSYE 129  
 Db 1269 KTDGTVNTKDLAVDTTEGPFAMKTNNTTGLTREASVNFQV-----STDSNGNYA 1320  
 Qy 130 EFDVKNPGGYTLSDAYRYFALAREM-KGNSFTPKPKSVNSIFHTVMRNDGLSIISVNVV 188  
 Db 1321 QL---TPQGTKLSTDPVMTDITRASGVTGSGTLRARINNGVFA--QSKDVTIPSIAPN- 1374  
 Qy 189 CYDPSLNL-SISPYICGAGVDAIEFDVLHIKFAVQSKGLGIAYSIPSNISLIFASLYYH 247  
 Db 1375 ----SLITIGTMSKFKSGVSGFDT-----LGLLYS-PGQLSV-ASV--- 1409  
 Qy 248 KVMGNQFKNLNV 259  
 Db 1410 -TVNGDGK-INI 1419

RESULT 13  
 TI7515  
 hypothetical protein A25L - Chlorella virus PBCV-1  
 C;Species: Chlorella virus PBCV-1  
 C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C;Accession: TI7515  
 R;Graves, M.V.; Van Etten, J.L.  
 submitted to the EMBL Data Library, May 1999  
 A;Reference number: Z18806  
 A;Accession: TI7515  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-401 <GRA>  
 A;Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96393.1  
 A;Experimental source: specific host Chlorella strain NC64A  
 C;Genetics:  
 A;Gene: A25L

Query Match 7.2%; Score 107; DB 2; Length 401;  
 Best Local Similarity 18.4%; Pred. No. 1.9;  
 Matches 61; Conservative 56; Mismatches 140; Indels 74; Gaps 10;  
 Qy 9 IINTVLVCLSLPNISSSKAINNAKYYGLYISGQ-YKPSVSV----- 51  
 Db 9 ISNTWQVSNLIVSNIS---AFNVITANSFQALSMGTGRYRPGISIERAGNLIVNSLSTSTA 65  
 Qy 52 -----PSNFSVKET-----NVITKNIALKKDVDSIETKTDASVGIS 88  
 Db 66 NILQVLTSNTGQVNITNTMFGNISMSNIIVTGVTGNTLTMGNVIANAVTTSVSTITGNA 125

Qy 89 NPENFTIPTAVFQDNSVNFNGTIGYTFAGTRVEIGSYEEDFVRNPGGYTLSDAYRYF 148  
 Db 126 TFGNLTVDVAVIRGNITANN-----ISTGSVIYNGAFITGNLTSSGNANISIAIRMF 179  
 Qy 149 ALAREMK-----GNSFTPKPKSVNSIFHTVMRNDGLSIISVNVVVCYDFSLNNLSIS 200  
 Db 180 GTNSAINVAGFTFGVGNVTNTNATIANLTNTLLSIGNLSTTG-----NFTVNSLUSVG 232  
 Qy 201 PYICGAGVDAIEFDVLHIKFAVQSKGLGIAYSIPSNISLIFASLYYHKVMGN-----QPKN 256  
 Db 233 NLFANGSTIQLSNLANIYQVSNLIYSGNVVV-----SNLSA-NLTLSYQVNIQTSTQLQSGT 286  
 Qy 257 LNVQHVLAELASIPKITSAVATLNIYVPGGEI 287  
 Db 287 MVMVNNILSAGFRRLSISSTAPVAGWTMGRV 317

RESULT 14  
 AC1347  
 probable peptidoglycan bound protein (LPXTG motif) lmo2179 [imported] - Listeria monocyt.  
 C;Species: Listeria monocytogenes  
 C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
 C;Accession: AC1347  
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussauguet, O.; Entian, K.D.; Feihl, H.  
 D.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A;Authors: Krest, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
 ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
 A;Title: Comparative genomics of Listeria species.  
 A;Reference number: AB1077; MUID:21537279; PMID:11679669  
 A;Accession: AC1347  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-1152 <GLA>  
 A;Cross-references: GB:NC\_003210; PIDN:CAD00257.1; PID:g16411649; GSPDB:GN00177  
 A;Experimental source: strain EGD-e  
 C;Genetics:  
 A;Gene: lmo2179

Query Match 7.1%; Score 105.5; DB 2; Length 1152;  
 Best Local Similarity 20.5%; Pred. No. 9.4;  
 Matches 51; Conservative 49; Mismatches 106; Indels 43; Gaps 8;  
 Qy 22 NISSSKAINNAKYYGLYISGQYKPSVFSNFSVKET-----NVITKN-----L 67  
 Db 610 NLESNKASISRLK-----GPFAMVLLPSGVKMKNTQNGPKVTVLNEYNQAGRQL 661  
 Qy 68 IALKKO-----VDSIETKTDASVGISNPSNFTIPTAVFQDNSVNFNGTIGYTFAGETR 121  
 Db 662 LKVKWDKKTLLPAEKLTAKNVSVAKDTPSNMTVEMFGFLQDTPNFVPEVSGTPTISDTK 721  
 Qy 122 VEIE-----GSYEEDFVRNPGGYTLSDAYRYFALAREMKGNSFTPKPKSVNSIFHTV- 173  
 Db 722 MEIDSDINQNGNSESRITSGNHYILNTS-NHLKISKAKGNRDKKEYSGLANATTNSIV 780  
 Qy 174 -----MENDGLSIISVNVVVCYDFSLNNLSISPICGAGVDAIEFDVLHIKFAVQSKL 228  
 Db 781 SYQLSLENDSEKIANMVLMDPLPSENDLGTITNSERGSKFN-LALTKAVEVPKEWKDKV 839  
 Qy 229 GIAYSLPSN 237  
 Db 840 EVIYSTAKN 848

RESULT 15  
 G86797  
 prophage p13 protein 09 [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
 C;Species: Lactococcus lactis subsp. lactis  
 C;Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C;Accession: G86797  
 R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich



[illegible]





Db 114 GVASTKDNVVTIDEINRAINSACAKAGLDNDKXHLHALPYRTLDKQEV--NDPLGMS-- 169  
Qy 118 EGTRVEIGSYEEFDVKNPGGTYLSDAYRYFALAREMKGNSTTPKEKVSNIFFHTWRND 177  
Db 170 -GTRLEV-----FTIHV-----YTEKONNIEN--LEKIMIQS 197  
Qy 178 GLSIISVIVN-----VCYD-----FSLNNLSISPIYCGGAGV 209  
Db 198 GVEIENIVNSAASIAIATLSNDERELGVACVDMGGETCNLTYSNSIRYKYLPGVSHH 257  
Qy 210 DAIEFDVHLHKFYAQSKLGIAYSLSFNSISLFASLYYHKVMGNQFNKLVNQ----- 260  
Db 258 LTTDLSHMLNTPPVAEKKIKY---GDLSPFG-----GEETPSQNVQIPTTGSDGH 306  
Qy 261 --HVAELASIPKI 271  
Db 307 ESHIVPLSEIQPI 319  
  
RESULT 27  
C81265  
probable lipoprotein Cj1678 [imported] - Campylobacter jejuni (strain NCTC 11168)  
C:Species: Campylobacter jejuni  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
C:Accession: C81265  
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel  
Nature 403, 665-668, 2000  
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
A:Reference number: A81250; MUID:20150912; PMID:10688204  
A:Accession: C81265  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-928 <PAR>  
A:Cross-references: GB:AL139079; GB:AL111168; NID:G6968971; PIDN:CAB73665.1; PID:G696909  
A:Experimental source: serotype O2, strain NCTC 11168  
C:Genetics:  
A:Gene: Cj1678  
  
Query Match 6.6%; Score 98.5; DB 2; Length 928;  
Best Local Similarity 24.8%; Pred. No. 23;  
Matches 79; Conservative 38; Mismatches 112; Indels 89; Gaps 15;  
  
Qy 2 NKKLFTIINTVLVCLLSLPNLSKAINNAKKYGLYI--SGQYKPSVSVFNSFV-- 57  
Db 306 NTKLEGNIINT-----GNASIGSDIKIEGAKVEGLVNGQSGISGVSVGSSIDS 359  
Qy 58 --KETNVITKNLIALKKD--VDSIETKTDSVGIS----NPSNFTIPYTAVFQDQNSVNFN 109  
Db 360 ITNEGNGAISGITYVKDKSLDSITNTSTSGISGISTNNSDNKLEIS-----NSGNIG 414  
Qy 110 GTIGYTFAGTRVEIGSYEEFDVKNPGGTYLSDAYRYFALAREMKGNSTTPKEKVSNSI 169  
Db 415 G-----KISTGS-ADMVINSNGGTISGGI-----SSSGSGSTSISSNSQ 453  
Qy 170 FHTVNRNDGLSI-----ISVIVNVCYDFSLNNLSISPIYCGGAGVDAIEFFDVLHKF 222  
Db 454 GSTI--NNGITVSGSAQVEISNQSGVKDENGNTVTNN-----GSG----- 492  
Qy 223 AVQSKLGIAYSLSFNSISLFASLYYHKVMGNQFNKLVNQHVA-----ELASIPKITS 273  
Db 493 -----SVGIKDWLVSTDKNTGKLTNTVIGGSRAFNKVENITVDQNSVDLEELNDINNIIIS 548  
Qy 274 AVATLNIGYFG-----GEI 287  
Db 549 GVNQNNIGNIGTNGSGEI 566  
  
RESULT 28  
D81411  
probable lipoprotein Cj0629 [imported] - Campylobacter jejuni (strain NCTC 11168)  
C:Species: Campylobacter jejuni  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002

C:Accession: D81411  
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel  
Nature 403, 665-668, 2000  
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
A:Reference number: A81250; MUID:20150912; PMID:10688204  
A:Accession: D81411  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-978 <PAR>  
A:Cross-references: GB:AL139079; GB:AL111168; NID:G6967817; PIDN:CAB75265.1; PID:G696809  
A:Experimental source: serotype O2, strain NCTC 11168  
C:Genetics:  
A:Gene: Cj0629  
  
Query Match 6.6%; Score 98.5; DB 2; Length 978;  
Best Local Similarity 24.8%; Pred. No. 25;  
Matches 79; Conservative 38; Mismatches 112; Indels 89; Gaps 15;  
  
Qy 2 NKKLFTIINTVLVCLLSLPNLSKAINNAKKYGLYI--SGQYKPSVSVFNSFV-- 57  
Db 356 NTKLEGNIINT-----GNASIGSDIKIEGAKVEGLVNGQSGISGVSVGSSIDS 409  
Qy 58 --KETNVITKNLIALKKD--VDSIETKTDSVGIS----NPSNFTIPYTAVFQDQNSVNFN 109  
Db 410 ITNEGNGAISGITYVKDKSLDSITNTSTSGISGISTNNSDNKLEIS-----NSGNIG 464  
Qy 110 GTIGYTFAGTRVEIGSYEEFDVKNPGGTYLSDAYRYFALAREMKGNSTTPKEKVSNSI 169  
Db 465 G-----KISTGS-ADMVINSNGGTISGGI-----SSSGSGSTSISSNSQ 503  
Qy 170 FHTVNRNDGLSI-----ISVIVNVCYDFSLNNLSISPIYCGGAGVDAIEFFDVLHKF 222  
Db 504 GSTI--NNGITVSGSAQVEISNQSGVKDENGNTVTNN-----GSG----- 542  
Qy 223 AVQSKLGIAYSLSFNSISLFASLYYHKVMGNQFNKLVNQHVA-----ELASIPKITS 273  
Db 543 -----SVGIKDWLVSTDKNTGKLTNTVIGGSRAFNKVENITVDQNSVDLEELNDINNIIIS 598  
Qy 274 AVATLNIGYFG-----GEI 287  
Db 599 GVNQNNIGNIGTNGSGEI 616  
  
RESULT 29  
A82615  
surface protein XF1981 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: A82615  
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717; PMID:10910347  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: A82615  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1190 <SIM>  
A:Cross-references: GB:AE004017; GB:AE003849; NID:G9107083; PIDN:AAF84783.1; GSPDB:GN001  
A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrez, H.  
as-Neto, E.; Docena, C.; El-Dorri, H.; Facinani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig  
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tsuchako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z.



716ZY : 010951A

**A;Gene: A29**

```

Query Match          6.6%; Score 98; DB 2; Length 820; .
Best Local Similarity 22.4%; Pred.No. 22;
Matches 66; Conservative 36; Mismatches 120; Indels 72; Gaps 12;

QY      10  INTVLVCLLSLPNITSSKA-----INNNAKKYGLYISQYKPSVSFNSFSVKE- 59
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db       284  VSNLITSNVSASNASVSTLTFTIGNLTNNTVATISNLFVTGTIAGNTTFSFSVSLL 343
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY      60  -TWVITKQLIALKKVDISFTKTDASVGISNPSNFTTIPYTAVFQDNSVNFNGTIGYTFAE 118
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db       344  VANISGNN-----SIPFTFP-----NVLTSTNITDGNVAVTGN----- 379
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY      119  GTRVEIGSYEEDVKVPGGYTLSDAYRYPALAREMKGNSFTPKPKVNSIFHTVMRNDG 178
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db       380  -----GPTLN-----VTSLRANVLTTRGDVSNLTTDVINVTN 413
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY      179  LSIIS--VIVNVCVDFSLNNLSISPYICGGAGVDAIEFFDLVHLHKFAYQSKLGIAYSLP- 235
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db       414  VFISSANILANNVITVQMNIS--NVNITGNA-----EAFDY--SANSVLMSTLSVGGILSL 466
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY      236  SNISLPASLYHKVGNQFKNL-NVQHVAAELASPKLTSATATLNIGYFGGEIG 288
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db       467  GNVNTNDISFGNIVTGNFFANFANVGNVQGNANVITWSTLTLSNSAQIGNISG 520
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 33
T30856
protein F2 - Streptococcus pyogenes
C:Species: Streptococcus pyogenes
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C:Accession: T30856
R:Jaife, J.; Natanson-Yaron, S.; Caparon, M.G.; Hanski, E.
Mol. Microbiol. 21, 373-384, 1996
A:Title: Protein F2, a novel fibronectin-binding protein from Streptococcus py
A:Reference number: Z20907; MUID:97011581; PMID:8858591
A:Accession: T30856
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1039 <JAF>
A:Cross-references: EMBL:U31980; NID:g1654115; PID:g1654116; PIDN:AAC44522.1
C:Genetics:
A:Note: PRTF2

```

Query Match	6.6%	Score 98	DB 2	Length 1039;
Best Local Similarity	22.4%	Pred. No. 29;		
Matches	81	Conservative 49	Mismatches 128	Indels 104; Gaps 20
Qy	3	NKLKFTIINTV-----LVCLSLPNLNISSKAINNNAKKYGLYISQO---YKPSVSV- 51-		
Db	123	HQLTYTFNTVIAGLDKVQLSAELS-LPLENEKVLVENTNISDFKSTIGGQEITYKTGVNVL 181		
Qy	52	FSNFVKETNVITKNLIKALDKVDISIETKTDASVG--ISGNPSNFTPIPTAV----- 100		
Db	182	YGNESTRKSNVITNGLSNVGGSIESYNTEGTGEFVVYVYVVPNRTPIPYAVLNWGFAXRT 241		
Qy	101	----FDNSVNFNNGTY-GYTFAE-----GTRVEIGSYBEFDVKNPGCYTILSDAY 145		
Db	242	AQGENDNLVSQAQLTGDIYEVPHNRYLRPTSYGVDISRNLRKDLKLEAKLPQGSTQGANK 301		
Qy	146	RY-FALAREMKGSNF-----TPKEKYSNISFIHTVMRNDGLSIIISIVINVCYDRL 194		
Db	302	RLRIDFGENLGKGAFVVKVTGADQSGKELIVQS--HLSSFNNMWSYKTLRPNSHVSEF- 358		
Qy	195	NNLSISPYICGGAGVDATIEFFDVLIHKIFAYOKSLGIAYSLSPNISLFASYHYHXYM--- 250		
Db	359	NEIALSP-SKGSMSGTSEF-----TKPSPITV---ANLKRVAAQLRFKKYSTDNV 402		
Qy	251	-----GNQFK---NLNVQ---HVAELAS-----IPK-----ITSAVATL 278		
Db	403	PLPEAAPELRSSNNGSQKLEASSNTQGBIHFKDLTSGYTDIYETKAPKGYQQVTEKLTAV 462		
Qy	279	NI 280		

Db 463 TV 464

RESULT 34  
F87662

TonB-dependent receptor, probable [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: F87662  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heide  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fra  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: F87662  
A:Status: preliminary  
A:Accession: F87662  
A:Molecule type: DNA  
A:Residues: 1-722 <STO>  
A:Cross-references: GB:AE005673; NID:g13425036; PIDN:AAK35298.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC3336

Query Match 6.5%; Score 97.5; DB 2; Length 722;  
Best Local Similarity 22.4%; Pred. No. 20;  
Matches 81; Conservative 44; Mismatches 128; Indels 109; Gaps 19;

Qy 20 LPNISS---SKAINNNAKTYGLYIS-GQYKPSVSFNS--FSVKETNVITKNL--IALK 71  
Db 306 IPNFIALLLNRPNTSFTPGGLYDRYDQRAKTVALFTNETHWVTDAPETIAGLRVTVEK 365  
Qy 72 KDVSIDSTKTDASVGINSPNFTPIPTAVFQDINSVNFNGTIGYTFASGTRV----- 122  
Db 366 KDLATFQTNDSGGVCGC-----TALTAPAGQARMAGIVG-AAAVGTIVGNLCLEPAN 415  
Qy 123 ---EIEGSEEFVKNFGGYTLSDAYR-----YFALAREMKGNSFT----- 160  
Db 416 PLFNGRGTQERTDKWSG-TIKASYRFPSEVFTYASFAKYGKGGGNLDRQTQSSNGLPS 474  
Qy 161 -----PKEKVSN---SIFITVMRNDGLSISIVNVYCYDFSLNNL----- 197  
Db 475 GSGSVTPYIDTSFPAEFVDSVYELGMKNTLFNRSVLFNVSFLFQOKFTDFQLNLTGTSFV 534  
Qy 198 -SISPYTCGGAGVDAIEFFD-----VLHIKAY-OSKLG-----TAYSLPSNISL 240  
Db 535 RSIPSVTSKGVADFMFTFVRGLVIOSGFTYADTKYQKQIPNDPGNALALPGSNLSL 594  
Qy 241 -----FASLYYHKVMGNQPK-----NLNVQVHAELAS-----IPKITSAVATLNIGYFGE 286  
Db 595 APKYSGSASVTYEAHSVGNLKNARNICAKYSSEYNTGSDLPFPKLOKAFVTVN-----GR 649  
Qy 287 IG 288  
Db 650 VG 651

RESULT 35  
H87341

OmpA family protein [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: H87341  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heide  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fra  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: H87341  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-407 <STO>

A;Cross-references: GB:AE005673; NID:g13421980; PIDN:AAK22732.1; GSPDB:GN00148  
C;Genetics:  
A;Gene: CC0747

Query Match 6.5%; Score 97; DB 2; Length 407;  
Best Local Similarity 22.0%; Pred. No. 11;  
Matches 51; Conservative 24; Mismatches 83; Indels 74; Gaps 7;

Qy 107 NFN-----GTIGYTFASGTRVEIEGSYEEDVKVPGGYTLDAYRYFALAREMKNS 158  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
Db 54 NFNADKDWAFAAGRLGYRFPNRVEAYGYRPSDLQGVRGAGLSGAQPGLCTAGVGRS 113  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
Qy 159 FTPKEKVSNIPIHTWRNDGLSIISIVNCYDFSLNNLSISPVCYGAGVDIAE----- 213  
:|| : || : || : || : || : || : || : || : ||| :  
Db 114 ASP-----GCCEPENGELKASTLMANLIFDMG-GDSRLSPFGVGAGTAWHNVKYG 163  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
Qy 214 -----PPDLVLHIKFAYOSKLGIAYSLSNLSLFASLYYHKVMGNQPKNLN 258  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
Db 164 QLSGVPAGAIIYQNTGFDDVDQAFAMQGLLGVAWNFADNWSM-----DLTRYLR-- 214  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
Qy 259 VQRVHAELASIPKTSAVATLNIGYFGGEI-----CARLTFF 293  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
Db 215 -----SKLDWGSVTQNAGPAGSITDVGTFSGRYKDTSTVLGLRYTF 256  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

RESULT 36  
S54039  
hypothetical protein YDR055w - yeast (*Saccharomyces cerevisiae*)  
N;Alternate names: hypothetical protein D4214; hypothetical protein YBRB444; hypothetical  
C;Species: *Saccharomyces cerevisiae*  
C;Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 19-Apr-2002  
C;Accession: S54039; S58835; S61744; S67871  
R;Hunt, S.; Bowman, S.; Harris, D.  
submitted to the EMBL Data Library, May 1995  
A;Reference number: S54031  
A;Accession: S54039  
A;Molecule type: DNA  
A;Residues: 1-444 <HUN>  
A;Cross-references: EMBL:Z49209; NID:g798897; PIDN:CAA89084.1; PID:g798906  
R;Brandt, P.; Otto, B.; Ramlow, S.; Blocker, H.  
submitted to the EMBL Data Library, January 1995  
A;Reference number: S58832  
A;Accession: S58835  
A;Molecule type: DNA  
A;Residues: 1-444 <BRA>  
A;Cross-references: EMBL:X84162; NID:g706817; PIDN:CAA58971.1; PID:g706821  
R;Brandt, P.; Ramlow, S.; Otto, B.; Blocker, H.  
Yeast 12, 85-90, 1996  
A;Title: Nucleotide sequence analysis of a 32,500 bp region of the right arm of Sacchar  
A;Reference number: S61741; MUID:96381250; PMID:8789263  
A;Accession: S61744  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-444 <BRW>  
A;Cross-references: EMBL:X84162; NID:g706817; PIDN:CAA58971.1; PID:g706821  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1995  
R;Blocker, H.; Brandt, P.  
submitted to the Protein Sequence Database, July 1996  
A;Reference number: S67587  
A;Accession: S67871  
A;Molecule type: DNA  
A;Residues: 1-444 <BLO>  
A;Cross-references: EMBL:Z74351; NID:g1431498; PIDN:CAA98873.1; PID:g1431499; MIPS:YDR05  
A;Experimental source: strain S288C  
C;Genetics:  
A;Gene: SGD:PST1  
A;Cross-references: SGD:S0002462  
A;Map position: 4R  
C;Keywords: transmembrane protein  
P;5-21/Domain: transmembrane #status predicted <TM>

Query Match 6.5%; Score 97; DB 2; Length 444;  
Best Local Similarity 22.5%; Pred. No. 12;



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 10:08:15 ; Search time 23 Seconds  
(without alignments)  
528.372 Million cell updates/sec

Title: US-10-062-624-40

Perfect score: 1496

Sequence: 1 MNKRLFTIINTVLCLSL.....AVATLNGYFGBIGARLTF 293

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	294	19.7	282	1 MSP4 ANAMA	Q07408 anaplasma m
2	101.5	6.8	527	1 PPCK_FUSNN	Q0812 fusobacteri
3	99	6.6	450	1 SAG1 YEAST	P20840 saccharomyc
4	98.5	6.6	492	1 FTSA_HELPY	O25629 helicobacte
5	97	6.5	444	1 PST1 YEAST	Q12355 saccharomyc
6	96	6.4	380	1 RMAR_WILMR	P47906 williopsis
7	94	6.3	483	1 SYD_WYCCA	Q48979 mycoplasma
8	94	6.3	2199	1 DPOE_SCHPO	P87154 schizosacch
9	93.5	6.2	296	1 KHRI YEAST	P22313 saccharomyc
10	93.5	6.2	1116	1 YK54 AQUAE	O67838 aquifex ae
11	93	6.2	738	1 Y013_BPL2	P42548 bacterioph
12	92.5	6.2	458	1 NUON_RICPR	Q92d13 rickettsia
13	92.5	6.2	774	1 YNA6 YEAST	Q04233 saccharomyc
14	92	6.1	463	1 DNAA_RICPR	Q59758 rickettsia
15	92	6.1	468	1 MURC_BORBU	O51757 borrelia bu
16	92	6.1	1513	1 STU1 YEAST	P38198 saccharomyc
17	91.5	6.1	439	1 Y441_BUCAI	P57516 buchnera ap
18	91.5	6.1	470	1 PUR1_METJA	Q57657 methanococc
19	91	6.1	1024	1 Y075_MYCGE	P47321 mycoplasma
20	90.5	6.0	1286	1 AIDA_ECOLI	Q03155 escherichia
21	90.5	6.0	1682	1 MSP1_PLAF3	P19598 plasmodium
22	90	6.0	463	1 DNAA_RICCN	Q92h56 rickettsia
23	90	6.0	956	1 HLYA_ACTSU	Q00951 actinobacil
24	89.5	6.0	385	1 DP3B_BORBU	P33761 borrelia bu
25	89.5	6.0	703	1 CDGT_BACS2	P31746 bacillus sp
26	89.5	6.0	711	1 PRE3 YEAST	Q08905 saccharomyc
27	89.5	6.0	740	1 BXC1_CLOBO	Q95242 sus scrofa
28	89.5	6.0	1290	1 BXC1_CLOBO	P18640 clostridium
29	89.5	6.0	1630	1 MSP1_PLAPK	P04932 plasmodium
30	89.5	6.0	1639	1 MSP1_PLAPW	P04933 plasmodium
31	89	5.9	444	1 SLAP_LACAC	P35829 lactobacill
32	88.5	5.9	237	1 OPAJ_NEIGO	Q04882 neisseria g
33	88.5	5.9	405	1 ACKA_BUCAI	P57272 buchnera ap

34	88.5	5.9	760	1 PBPB_BUCAI	P57286 b penicilli
35	88	5.9	238	1 QP66_NEIGO	Q05033 neisseria g
36	88	5.9	334	1 SD55_CAEEL	Q09572 caenorhabdi
37	88	5.9	613	1 CGAA_CLOBI	Q45882 clostridium
38	88	5.9	956	1 RT2A_ACTPL	P15377 actinobacil
39	88	5.9	1273	1 MY53_YEAST	P36006 saccharomyc
40	88	5.9	1476	1 GTFB_STRMU	P08987 streptococc
41	87.5	5.8	273	1 ROCI_NICSY	Q08935 nicotiana s
42	87.5	5.8	364	1 PGLR_COCCA	P26215 cochllobolu
43	87.5	5.8	445	1 WAPA_STRMU	P11000 streptococc
44	87.5	5.8	602	1 PEHX_ERWCH	P15922 erwinia chr
45	87.5	5.8	1032	1 TLRO_MOUSE	P58682 mus musculus

ALIGNMENTS

RESULT 1	MSP4 ANAMA	STANDARD;	PRT;	282 AA.
ID	MSP4 ANAMA			
AC	Q07408;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	01-NOV-1995 (Rel. 32, Last annotation update)			
DE	Major surface antigen 4 precursor.			
GN	MSP4.			
OS	Anaplasma marginale.			
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;			
OC	Anaplasmataceae; Anaplasma.			
OX	NCBI_TaxID=770;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 30-53.			
RX	MEDLINE=94124017; PubMed=8294020;			
RA	Oberle S.M., Barbet A.F.;			
RT	"Derivation of the complete msp4 gene sequence of Anaplasma marginale without cloning."			
RL	Gene 136:291-294 (1993).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.iesb-sib.ch/announce/ or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL; L01987; AAC38877.1; --			
DR	InterPro; IPR002566; Surface_Ag_msp4.			
DR	Pfam; PF01617; Surface_Ag_2; 1.			
KW	Antigen; Signal.			
FT	SIGNAL			
FT	CHAIN 30 282 MAJOR SURFACE ANTIGEN 4.			
SQ	SEQUENCE 282 AA; 239997 MW; 3088D31085A8834B CRC64;			
	Query Match 19.7%; Score 294; DB 1; Length 282;			
	Best Local Similarity 26.8%; Pred. No. 2.4e-15;			
	Matches 82; Conservative 57; Mismatches 129; Indels 38; Gaps 9;			
Qy	1 MNKRLFT-----INTVLVCLLSLPNTSSSAINNNAKYGL-----YISQYKPSVSVP 52			
Db	1 MNYRELFTGLSAATVACACSLLVSGAVVAPSPMSHEVASEGGVMGSGFYVGAAYSPAPFSV 60			
Qy	53 SNFSVKETNVTIKNLIALKXDVDSIETKTASVIGSNPSNFT-IPYTAVFQDNV-NFNG 110			
Db	61 TSFDMRESSKETSVRGYDKSI-----ATIDVSPANFSGSYTAFSKNLTISFDG 112			
Qy	111 TIGYTFAGTRVEIEGSEYBEFVKNPQGYTLSDARYFALARE---MKGNSFTPEKVSVN 167			
Db	113 AVGYSLG-GARVELEASYREFATLADQYAKSGAESLAATRDANITETNVF-----163			
Qy	168 SIHFTVNRNDGLSIISVINVCYDFSLNNISIPYICGGAGVDAIEFDVLHIFAYOSK 227			
Db	164 ----VVKIDEINTNSVNLNGCYDVLTLDPVSPVCGAGISGFVDISKQVTTKLAYRGK 218			



```
CC -----
DR EMBL; M28164; AAA34417.1; -
DR EMBL; X16861; CAA34752.1; -
DR EMBL; X87611; CAA60926.1; -
DR EMBL; Z49504; CAA89526.1; -
DR PIR; S22835; S22835.
DR SGD; S0003764; SAG1.
KW Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat.
FT SIGNAL 1 19
FT CHAIN 20 650
FT DOMAIN 278 350
FT
FT DOMAIN 339 423
FT REPEAT 339 378
FT REPEAT 384 423
FT DISULFID 97 114
FT DISULFID 202 300
FT CARBOHYD 79 79
FT CARBOHYD 109 109
FT CARBOHYD 135 135
FT CARBOHYD 248 248
FT CARBOHYD 282 282
FT CARBOHYD 289 289
FT CARBOHYD 299 299
FT CARBOHYD 303 303
FT CARBOHYD 306 306
FT CARBOHYD 307 307
FT CARBOHYD 308 308
FT CARBOHYD 311 311
FT CARBOHYD 314 314
FT CARBOHYD 315 315
FT CARBOHYD 316 316
FT CARBOHYD 329 329
FT CARBOHYD 331 331
FT CARBOHYD 334 334
FT CARBOHYD 335 335
FT CARBOHYD 338 338
FT CARBOHYD 339 339
FT CARBOHYD 340 340
FT CARBOHYD 341 341
FT CARBOHYD 342 342
FT CARBOHYD 345 345
FT CARBOHYD 346 346
FT CARBOHYD 349 349
FT CARBOHYD 350 350
FT CARBOHYD 364 364
FT CARBOHYD 402 402
FT CARBOHYD 460 460
FT CARBOHYD 485 485
FT CARBOHYD 501 501
FT CARBOHYD 614 614
FT CONFLICT 449 449
FT CONFLICT 556 556
FT CONFLICT 581 581
SQ SEQUENCE 650 AA; 70339 MW; 8BBF7A1C44C3C2B CRC64;

Query Match
Best Local Similarity 20.9%; Score 99; DB 1; Length 650;
Matches 65; Conservative 46; Mismatches 120; Indels 80; Gaps 13;

Qy 8 TIINVLVCLSLPNISSKAINNAKKYGLYISGGYKPSVSVFNSFKETNV-----62
Db 345 TSINT---SAYSTGSIETVETGNTTSEVISHVVTSTKLSPTATTSITIAQTSIYSDS 401
Qy 63 -----ITKNLIALKDDVDSIETKTDASV-----GISNP--SNFT-----94
Db 402 NITVGTDIHTTSEVIS---DVETISRETASTVAAPTSTTGWTGAMNTIYSQFTSSSPAT 458
Qy 95 -----IPYTAFO--DNSV-----NPNFGTIGY-----TEAGTRVEIEGSYEEDVKN 135
Db 459 INSTPISSSAVFTSDASIVNVHTENITNTAAPPSEPTFNATRNLSNFCSCSKQPS 518
Qy 136 PGGYTLSDARYFALAREMKGNSFTPKPKVNSIFHTVMRNDG-----LSIISVIVNVC 189
```

```
Db 519 PSSYTSPLVSVSVKSTLLSTFTPSVPTNTYIKT--KNTGYFEHTALTTSVGLNSP 576
Qy 190 YD-----FSLNLSISPIYICGGAGVDAIE--FFDVLHLIKPAYQSKLGIAYSLP 235
Db 577 SETAVSQGTAKIDTFLVSSLIAYPSSASGSLGSIQONFTSTLSIMISTYEGKASIFFSAE 636
Qy 236 SNISLFLASLYY 246
Db 637 LGSIIIFLLSY 647

RESULT 4
FTSA_HELPY STANDARD; PRT; 492 AA.
AC O25629;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell division protein ftSA.
GN FTSA OR HP0978.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RP [1]
SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RC MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G., B.A.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty S.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RL pylori."
RC Nature 388:539-547(1997).
CC -!- FUNCTION: THIS PROTEIN MAY BE INVOLVED IN ANOMALOUS FILAMENT
CC GROWTH (BY SIMILARITY). MAY BE A COMPONENT OF THE SEPTUM. IT
CC MAY INTERACT WITH FTSZ (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE FTSA/MREB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF000606; AAD08024.1; -
DR TIGR; HP0978; -
DR InterPro; IPR003494; FtSA.
DR Pfam; PF02491; FtSA; 2.
DR TIGRFAMs; TIGR01174; ftSA; 1.
KW Cell division; Cell shape; Complete proteome.
SQ SEQUENCE 492 AA; 54492 MW; FA237B467BE39B4A CRC64;

Query Match
Best Local Similarity 6.6%; Score 98.5; DB 1; Length 492;
Matches 58; Conservative 46; Mismatches 102; Indels 107; Gaps 14;

Qy 19 SLPNISS--KAINNAKKYGLYIS-----GQYKPSVSVFNSFKETNV-----63
Db 54 SLAHASNAIKVINSKAKMAGLNADENRNNPMHFGEVHPKTKAIVSFGAYTESIRDVT 113
Qy 64 -----TKNLIALKDDVDSIETKTDASVGSINPSNF--TIPYTAVFQDINSVFNFGTGYFA 117
Db 114 GVASTKDNVVTIDEINRAINSACAKAGLDNDKHLHPYRFTLDKQEV--NDPLGMS-- 169
```

QY 118 EGTRVEIEGSEEDVKNPGGYTLDAYRYFALAREMKNSFTPKKVSNSIFHTVMRND 177  
 Db 170 -GTRLEV-----FIHIV-----YTKQNNIEN--LEKIMIQS 197  
 QY 178 GLSIISVIVN-----VCYD-----FSLNNLSISPIYCGGAGV 209  
 Db 198 GVEIENIVNSVAASIALNSDERELGVACVDMGGETCNLTYSNSIRYKILPVGSHH 257  
 QY 210 DAIEFFDLHTKFAVQSKLGIAYSPLSNISLFASYLHYHKWGNQPKNLNVQ----- 260  
 Db 258 LTTDLSHLNTFPFYVAEVEKIKY---GDLSPRG-----GEETPSQNVQIPTTGSDGH 306  
 QY 261 --HVAELASIPKI 271  
 Db 307 ESHIVPLSEIQTI 319

## RESULT 5

PS11 YEAST  
 ID PS11 YEAST STANDARD; PRT; 444 AA.  
 AC Q12355;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protoplast secreted protein 1 precursor.  
 GN PS11 OR YDR055W OR b4214 OR YD9609.09.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96381250; PubMed=8789263;  
 RA Brandt P., Ramlow S., Otto B., Bloecker H.;  
 RT "Nucleotide sequence analysis of a 32,500 bp region of the right arm  
 of Saccharomyces cerevisiae chromosome IV.";  
 RL Yeast 12:85-90(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RA Hunt S., Bowman S., Harris D., Barrell B.G., Rajandream M.A.;  
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP IDENTIFICATION.  
 RX MEDLINE=95251092; PubMed=10234784;  
 RA Pardo M., Monteoliva L., Pla J., Sanchez M., Gil C., Nombela C.;  
 RT "Two-dimensional analysis of proteins secreted by Saccharomyces  
 cerevisiae regenerating protoplasts: a novel approach to study the  
 cell wall.";  
 RL Yeast 15:459-472(1999).  
 RN [4]

## GPI-ANCHOR.

QY MEDLINE=20469049; PubMed=11016834;  
 RA Terashima H., Yabuki N., Arisawa M., Hamada K., Kitada K.;  
 RT "Up-regulation of genes encoding glycosylphosphatidylinositol  
 (GPI)-attached proteins in response to cell wall damage caused by  
 disruption of FKS1 in Saccharomyces cerevisiae.";  
 RL Mol. Gen. Genet. 264:64-74(2000).  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR AND  
 SECRETED BY REGENERATING PROTOPLASTS.  
 CC -1- SIMILARITY: BELONGS TO THE SPS2 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X84162; CAA58971.1; -.  
 DR EMBL; 274351; CAA98873.1; -.

DR EMBL; Z49209; CAA89084.1; -.  
 DR SGD; S0002462; EST1.  
 DR COMPLYEAST-2DPAGE; Q12355; -.  
 DR InterPro; IPR000494; EGFR\_L domain.  
 DR Pfam; PF01030; Recep\_L domain; 1.  
 KW Glycoprotein; Membrane; GPI-anchor; Signal.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 ? PROTOPLAST SECRETED PROTEIN 1.  
 FT PROPEP ? 444 REMOVED IN MATURE FORM (POTENTIAL).  
 FT DOMAIN 356 416 SER-RICH.  
 FT CARBOHYD 57 57 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 76 76 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 83 83 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 210 210 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 228 228 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 235 235 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 242 242 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 263 263 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 268 268 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 280 280 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 292 292 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 329 329 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 444 AA; 45776 MW; 230F60CAC5921A4 CRC64;

## Query Match 6.5%; Score 97; DB 1; Length 444;

Best Local Similarity 22.5%; Pred. No. 3.4;  
 Matches 64; Conservative 42; Mismatches 126; Indels 52; Gaps 10;  
 QY 17 LLSLPNISSKAINNAKKYGLYISQYKPSVSVFNSFKETNVI-----TKNLIALKK 72  
 Db 126 LITLPAISS---FTSNKSNANNIYISDTLSQSDVDFS--ALKKYNVFNANNKLTSTIKS 180  
 QY 73 DVDSIETKTDASVIGSNPSNFTIPTYTAVFQDNSVNFNGTIGTYFAEGTRVIEGSEYEDF 132  
 Db 181 PV---ETVSDLSQFSFNGNQTKITFDLWANNILSLTDVHSVSPANLNSSLGFINNS 237  
 QY 133 VKNFGGYTLSDAYRYFALARE--MKGNSFTPKKVSNSIFHTVMRNDGLSIISVIVN--- 187  
 Db 238 ISSLNFTKLTNTIGTFTSIVSNDYLNKLSFNSLTIGGAL--VVANNTGLQKIGLDNLTT 295  
 QY 188 -----VCYDFSLNNLSISPIYCGGAGVDAIEFPDVLHIFKPAYOSKLGIAYSLSNLSLF 241  
 Db 296 IGGTLEVVGNFTSLNLSLXSVKGGADVE-----SKSNFSN 333  
 QY 242 ASLYYHK---VMGNQF--KXNLNVQHVLAELASIPKITSATVLNI 280  
 Db 334 ALKALQKGGIKGESFVCKNGASSTSVKLSSTSKSQSQTAKV 377

## RESULT 6

EMAR WILMR  
 ID EMAR WILMR STANDARD; PRT; 380 AA.  
 AC P479067;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Mitochondrial ribosomal protein VAR1.  
 GN VAR1.  
 OS Williopsis mrakii (Yeast) (Hansenula mrakii).  
 OG Mitochondrion.  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Williopsis.  
 OX NCBI\_TaxID=4963;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CBS 1707;  
 RX MEDLINE=94287716; PubMed=8017108;  
 RA Triessi R., Sor F., Fukuhara H.;  
 RT "Genes of the linear mitochondrial DNA of Williopsis mrakii: coding  
 sequences for a maturase-like protein, a ribosomal protein VAR1



RT homologue, cytochrome oxidase subunit 2 and methionyl tRNA.";  
RL Yeast 10:391-398(1994).  
CC -!- FUNCTION: ESSENTIAL FOR MITOCHONDRIAL PROTEIN SYNTHESIS AND  
CC REQUIRED FOR THE MATURATION OF SMALL RIBOSOMAL SUBUNITS  
CC (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Mitochondrial.  
CC -!- SIMILARITY: BELONGS TO THE VARI FAMILY OF RIBOSOMAL PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X66594; CAA47158.1; -.  
KW Ribosomal protein; Mitochondrion.  
SQ SEQUENCE 380 AA; 44376 MW; CC3A100C5AD3A01A CRC64;  
  
Query Match 6.4%; Score 96; DB 1; Length 380;  
Best Local Similarity 22.9%; Pred. No. 3.3;  
Matches 58; Conservative 36; Mismatches 87; Indels 72; Gaps 14;  
  
Qy 3 NKLKPT-IINTVLVCLSLPLNMISSKAINNAKYYG----- 38  
Db 109 NRPKFKHTINTVYINFY--NDTNKMINNKHLYGSLINDINILGCFNYKNNHNF 166  
Qy 39 ---LYISQKPSVSVFNSFKV---ETNVTKNLIA--LKKVDSTIETKTDSVGSIN- 89  
Db 167 NIATVLSGLYKNNKVMIPN--KMKYNDVNDVFNSSISYDLKYGKGLAGKTYSKLLRDNI 225  
Qy 90 PSNETIP-----YTAVFQDQSVNFNGTIGYTFAGCTRVEIEGSEYEEFDVKNPGGYTLSDA 144  
Db 226 PMNNSLSIKNNYMTININNNIKNNMI-----SNNSLNIDIKYSPDKINTNELLVWK 280  
Qy 145 YRYFALAREMKN-----SFTPEK-----VNSIFHTVMRNDGLSI-----I 182  
Db 281 Y-LIGLSMLFPGKNIKAGVSRISKEILLFGLSNKLYR---KNSGLLVYKNNNNNTTKYL 336  
Qy 183 SVIVNVCYDFSLN 195  
Db 337 NFDININKYKLN 349  
  
RESULT 7  
SYD\_MYCCA  
ID SYD\_MYCCA STANDARD; PRT; 483 AA.  
AC Q48979;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Aspartyl-tRNA synthetase (EC 6.1.1.12) (Aspartate--tRNA ligase)  
DE (ASPRS) (fragment).  
GN ASPs.  
OS Mycoplasma capricolum.  
OC Bacteria; Firmicutes; Mollicutes; Entomoplasmatales;  
OC Entomoplasmataceae.  
OX NCBI\_TaxID=2095;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 27343 / KID;  
RX MEDLINE=96059641; PubMed=7476192;  
RA Bork P., Ouzounis C., Casati G., Schneider R., Sander C.,  
RA Dolan M., Gilbert W., Gillevet P.M.;  
RT "Exploring the Mycoplasma capricolum genome: a minimal cell reveals  
RT its physiology."  
RL Mol. Microbiol. 16:955-967(1995).  
CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate + tRNA(Asp) = AMP +  
CC diphosphate + L-aspartyl-tRNA(Asp).  
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z33048; CAA83720.1; -.  
DR HSP; P21889; LCOA.  
DR InterPro; IPR002106; AaRNA\_ligaseII.  
DR InterPro; IPR004115; GAD\_dom.  
DR InterPro; IPR004364; tRNA-synt\_2.  
DR Pfam; PF00152; tRNA-synt\_2; 2.  
DR Pfam; PF02938; GAD; 1.  
DR PROSITE; PS50862; AA tRNA LIGASE II; 1.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.  
FT NON\_TER 1  
SQ SEQUENCE 483 AA; 56298 MW; A03F5B0CEA31087E CRC64;  
  
Query Match 6.3%; Score 94; DB 1; Length 483;  
Best Local Similarity 21.2%; Pred. No. 6.3;  
Matches 53; Conservative 25; Mismatches 82; Indels 90; Gaps 9;  
  
Qy 31 NNAKYYGLYISQKPSVSVFNSFKVTKNTVITKNLIAKVDSTIETKTDSVGSINP 90  
Db 18 NNLKNEYVIEISGIVVVKRSVKELITGEIVVKDLINKS----- 60  
Qy 91 SNFTIPTAVFDQSVNFNGTIGYTFAGCTRVEIEGSEYEEFDVKNPGGYTLSDAYRFAL 150  
Db 61 -----ELTPFVLENDVNN-----EDTRL-----TYRYLDL 86  
Qy 151 AREMKNGSFTPEKSVNSIFHTVMRNDGLSIISVIVNVCYDFSLNLSI-SPYICGGAGV 209  
Db 87 RRQVMQNLIRAKIN-----IIRN-----YLTDLNLFLEVTPTPAKSTPE 128  
Qy 210 DAIEFFDVLHI---KPAYQSKGLIAYSLPSNISLFSAL-----YHKVMGNQFNLN 258  
Db 129 GARHFLVPSRLNKNKF-----YALPQSPQLFKLLMISGIDRYVQIVRCRDEDLR 179  
Qy 259 VQHVAEIASI 268  
Db 180 IDRQPEFTQL 189  
  
RESULT 8  
DPOB\_SCHPO  
ID DPOB\_SCHPO STANDARD; PRT; 2199 AA.  
AC P87154;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE DNA polymerase epsilon, catalytic subunit A (EC 2.7.7.7) (DNA  
DE polymerase II subunit A).  
GN CDC20 OR SPBC25H2.13C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,

RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer C., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas R., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,  
 RT "The genome sequence of Schizosaccharomyces pombe".  
 RL Nature 415:871-880(2002).  
 CC -!- FUNCTION: DNA POLYMERASE II PARTICIPATES IN CHROMOSOMAL DNA  
 CC REPLICATION (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
 CC + [DNA] (N).  
 CC -!- SUBUNIT: CONSISTS OF FIVE SUBUNITS (200 kDa, 80 kDa, 34 kDa, 30  
 CC kDa, AND 29 kDa) (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -!- DOMAIN: THE DNA POLYMERASE ACTIVITY DOMAIN RESIDES IN THE  
 CC N-TERMINAL HALF OF THE PROTEIN, WHILE THE C-TERMINUS IS NECESSARY  
 CC FOR COMPLEXING SUBUNITS B AND C (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. HIGH  
 CC SIMILARITY WITH MAMMALIAN DNA POLYMERASE EPSILON.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z95397; CAB08772.1; -;  
 CC InterPro; IPR002064; DNA pol B.  
 CC Pfam; PF00136; DNA pol B.1.  
 CC PRINTS; PR00106; DNAPOLB.  
 CC SMART; SM00486; POLBc.1.  
 CC Transferrase; DNA-directed DNA polymerase; DNA replication;  
 CC DNA-binding; Zinc-finger; Nuclear protein.  
 CC ZN FING 2069 2155 BY SIMILARITY.  
 FT SEQUENCE 2199 AA; 252885 MW; A80ASD0865EBBC3E CRC64;  
 SQ  
 Query Match 6.3%; Score 94; DB 1; Length 2199;  
 Best Local Similarity 21.2%; Pred. No. 41; Mismatches 116; Indels 88; Gaps 17;  
 Matches 68; Conservative 49;  
 QY 4 KLKFTIINTVLCLSLPNISSKAI-NNNAKYKLYISGQYKPSVSVFNSFKETN- 61  
 DB 1504 ELRSKFDNLIKESSTYPTLSCNVIFSGNERKAYLI-----DEKLQVFTKTKNS 1556  
 QY 62 --VITKNL-IALKKVDSDIETKTASGVISNPSNFTPTAV--FDNSWNFN----- 109  
 DB 1557 LLTISSLPFILKANVKQIE-----ELPYIMIPRLSENIQSLSWKHQIAT 1601  
 QY 110 GTIGYTFAGT-----RVETEG-----SYEFDKVKNPGGYTLSDARYFALAREMKNS 158  
 DB 1602 KMIQHLAIGSLWFHRIQSRSDIPLCNFESDDIQ-----YSIDVYV-----SRKLKEH 1652  
 QY 159 F-----TPK-----EKVNSGIFHTVMRNDGLSIIS--VIVNVCYDFSLNLSISPYIC 204  
 DB 1653 IILWNKGTPLDGGIEK--DSLQIATSPKDPLEVNPNPCAYSNACVDISLSNALCS-IL 1709  
 QY 205 GGAGVDATF-----FFDLVLIKFAYQSKLGIAYSPLNLSIFASLYHHKV 249  
 DB 1710 NSALINDIEGIDMAALNDNYMTAINDDLEELGIDHNTGLTSLPVLKALVKTWNAA 1769

QY 250 MGNOFKNLNVQHVLAELASIPK 270  
 DB 1770 SGNNLADLIIOHLARWISSK 1790  
 RESULT 9  
 ID KHR1 YEAST STANDARD; PRT; 296 AA.  
 AC P22313;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Killer toxin KHR precursor (Killer of heat resistant).  
 GN KHR1 OR KHR.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=18 RHO(-);  
 RX MEDLINE=91000581; PubMed=1368554;  
 RA Goto K., Iwatuki Y., Kitano K., Obata T., Hara S.;  
 RT "Cloning and nucleotide sequence of the KHR killer gene of  
 RT Saccharomyces cerevisiae".  
 RL Agric. Biol. Chem. 54:979-984(1990).  
 CC -!- FUNCTION: KILL SENSITIVE STRAINS OF YEAST.  
 CC -!- SIMILARITY: TO YEAST YER076C.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; D00905; BRA00751.1; -;  
 CC FIR; J00548; J00548.  
 CC SGR; L0000899; KHR1.  
 KW Toxin; Glycoprotein; Signal.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 296 KILLER TOXIN KHR.  
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 71 71 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 122 122 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 146 146 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 296 AA; 32781 MW; 965AD10BA9CDBF01 CRC64;  
 Query Match 6.2%; Score 93.5; DB 1; Length 296;  
 Best Local Similarity 21.1%; Pred. No. 3.8; Indels 51; Gaps 11;  
 Matches 60; Conservative 51; Mismatches 122;  
 QY 13 VLVCLLSLPNISSSKA-----INNNAKYKLYISG-QYKPSVSVFNSFKET 60  
 DB 6 ILFSIIAVIATATVASSDSIYLGHRVGQIDISLYRYDNGTWYPTFNEMLDLTGMN 65  
 QY 61 NVITKNLIALKVDSDIETKT-----DASGVISNPSNFTPTAVFQNSWNFN 109  
 DB 66 DLATNNATILKRDSSVSCVTETCYQYVDYHVDDEGVITIDISTYRIP---VEWDSGSAGN 122  
 QY 110 GTIGYTFAGT-----RVETEG-----SYEFDKVKNPGGYTLSDARYFALAREMKNSFTPKKVS 166  
 DB 123 ASYGVS-----KRDTKYETFCCKKIICGINVSGFCNAIDFAVHAFDGGSVINPVSGIT 175  
 QY 167 NSIFHTVMRND-----GLSIISVIVNVCYDFSLNLSISPYICGGAGVDATFDFDLVLIKF 222  
 DB 176 DRIEATKRDTECLGYELDHVRIDPAVDWS---ISISTWKQSSANCDTQASADSL--KC 230  
 QY 223 AYQSKLGIAYSPLNLSIFASLYHHKVMGNQFKNLVQHVLAEL 266  
 DB 231 AAKALESEHNHOK-----TAFCHLNDGGSGF-NLDIRLISELS 269

```
RESULT 10
YK54_AQUAE
ID YK54_AQUAE STANDARD; PRT; 1116 AA.
AC O67838;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AQ_2054.
GN AQ_2054.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Shead M.A., Keller M., AuJay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358 (1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L13696; AAA87969.1; -.
KW Hypothetical protein; Signal; Transmembrane.
FT SIGNAL 1 27
FT CHAIN 28 738
FT TRANSMEM 612 732
FT TRANSMEM 712 732
FT SEQUENCE 738 AA; 81312 MW; 87479A3267C9846 CRC64;
SQ
Query Match 6.2%; Score 93; DB 1; Length 738;
Best Local Similarity 22.2%; Pred. No. 13;
Matches 52; Conservative 39; Mismatches 81; Indels 62; Gaps 10;
Qy 5 LKFTIINTLVCLSL-----LPNISSSKAINNNAKKYGLYISQYKP-SYVSFVS-NFSV 57
Db 378 LNYSDVMVAVNDLLTPLEDFLNNVAVDAIDGDISIVITDNDGYSPTVGIYNVFVSFV 437
Qy 58 KETVITKNIALKKVDTSIETKTDASVGISNPSNFTIPYTAVFQDNSVNFNGTIGYTFA 117
Db 438 TNSNGQTSIIAPVHVVD-----IVNP-----VI 461
Qy 118 EGRVIEGYSYEE-FDVKN-PGGVTLSDAYRYPALAREMKNST-PKEKVS-----166
Db 462 NGVSDTVHISYDQTFNVTNWNVSLTSDNY-YTGLSISIKENTYTVNKNKLGTYKITVQA 520
Qy 167 -----NSIFHTWRNDGLSIIISVIIVCVYDSESLNLSISPYICCGAGVDAIE 213
Db 521 VDPSNGITLRTITVVDNGIGPVGNGINTTASINENITVEQIKAGLAADAID 574

RESULT 12
NUON_RICPR
ID NUON_RICPR STANDARD; PRT; 458 AA.
AC Q92D13;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NADH dehydrogenase I chain N (EC 1.6.5.3) (NADH-ubiquinone
DE oxidoreductase chain N).
GN NUON OR RP537.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sichertitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
```





OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C;  
RX MEDLINE=95105240; PubMed=7806575;  
RA Pasqualone D., Huffaker T.C.;  
RT "STU1, a suppressor of a beta-tubulin mutation, encodes a novel and  
RT essential component of the yeast mitotic spindle.";  
RL J. Cell Biol. 127:1973-1984(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C;  
RX MEDLINE=94378718; PubMed=8091857;  
RA Skala J., van Dyck L., Furnelle B., Goffeau A.;  
RT "The sequence of an 8.8 kb segment on the left arm of chromosome II  
RT from Saccharomyces cerevisiae reveals four new open reading frames  
RT including homologs of animal DNA polymerase alpha-primases and  
RT bacterial GTP cyclohydrolase II.";  
RL Yeast 10:S13-S24(1994).  
CC -!- FUNCTION: ESSENTIAL COMPONENT OF THE MITOTIC SPINDLE. MAY CROSS-  
CC LINK POLAR MICROTUBULES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z35655; CA84714.1; -;  
DR EMBL; Z35795; CA84854.1; -;  
DR EMBL; X74738; CA52760.1; -;  
DR PIR; S45768; S45768.  
DR SGD; S0000130; STU1.  
KW Mitosis; Microtubules.  
SQ SEQUENCE 1513 AA; 174176 MW; E7F64B5A5D03588D CRC64;

Query Match 6.1%; Score 92; DB 1; Length 1513;  
Best Local Similarity 18.7%; Pred. No. 37;  
Matches 65; Conservative 60; Mismatches 92; Indels 130; Gaps 17;

QY 1 MNKLKFTTINTVLVCLSLPNISSKAINNAKYYGLYISGOYKPSVSVSNFSVKET 60  
DB 641 LSKLKTNLRS- - - - -SEYSSKENEKR- - - - -ARHDSNMNSVSNSTKDN 681  
QY 61 NVITKLNIALKKDVDSIETKTDAVSGISNPSNFTPTPTAVFDQNSVNFNGTIGYTPA- 117  
DB 682 NNVT- - - - -KKKVSAPPSTATKVSENYTFD- - - - -DFPSNQIDLTDELSNSYNPL 730  
QY 118 - - - - -EGTRVEIEGS- - - - -YEFP-DVKNPGGYTLSDAIRYFALAREM- 154  
DB 731 IKKYMKNQNDVSMSSPSISLKGSLKGEYETLYKKFNDAFFA-QIKDALQY-LQKELL 786  
QY 155 - - - - -KGNSTPPK- - - - -EKNVSI- - - - -PHTVMND 177  
DB 787 LTSQSSSAPKPEFFPMIMKLRQIMKSPNDPKPFLSIEKFTNGVPLNYLIBELYSINSPD 846  
QY 178 GLSIIISVIVNCVDSFSLNLSISPYICGGAGVDAIEFFDLVHLIKPAYQSKGIAYSLSPN 237  
DB 847 YAEILKRNKNPKPELTNLIIT- - - - -IADLFN- - - - -ANNCPND 885  
QY 238 ISLFSALY- - - - -YH-KVMGNQKRLNVQHVLAELAS- - - - -IPKIT 272  
DB 886 FKLYMYKKTTFNFYFNKLLLEIFRNLNKHONTLRSGTNDLMPKIS 932

RESULT 17  
Y441\_BUCAI  
ID Y441\_BUCAI  
AC P57516;  
DT 16-OCT-2001 (Rel. 40, Created)  
PRT; 439 AA.

DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein BU441.  
GN BU441.  
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum  
OS symbiotic bacterium).  
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.  
OX NCBI\_TaxID=118099;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Tokyo 1998;  
RX MEDLINE=20445173; PubMed=10993077;  
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;  
RT "Genome sequence of the endocellular bacterial symbiont of aphids  
RT Buchnera sp. APS.";  
RL Nature 407:81-86(2000).  
CC -!- SIMILARITY: BELONGS TO THE UPF0004 FAMILY. STRONG, TO E.COLI YLEA.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AP001119; BABI3139.1; -;  
DR InterPro; IPR002792; TRAM.  
DR InterPro; IPR001861; UPF0004.  
DR Pfam; PF00919; UPF0004; 1.  
DR Pfam; PF01938; TRAM; 1.  
DR TIGRFAMs; TIGR00089; UPF0004; 1.  
DR PROSITE; PS01278; UPF0004; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 439 AA; 50464 MW; 287361B42483D610 CRC64;

Query Match 6.1%; Score 91.5; DB 1; Length 439;  
Best Local Similarity 20.3%; Pred. No. 8.7; Indels 105; Gaps 14;  
Matches 68; Conservative 44; Mismatches 44

QY 27 KAINNAKYYGLYISGOYKPSVSVSNFSVKETNVTIKNLIALKKDVDSIETKTDAVSG 86  
DB 67 KIKNNPKVITAVGCVATQSGKEIFKRANYVDIIFGTQTLRLPKMIDEVEKKRLSID 126  
QY 87 ISNP- - - - -SNFTIPYT- - - - -AVFQ- - - - -102  
DB 127 ISFPKLEKFKYFLAPKKKGYTADISIMEGCKYSCFVVPYTRGNEISRCPDDVLFSL 186  
QY 103 - - - - -DNSVN- - - - -ENGITGYTFAEGTRV- - - - -EIEGSYE-EFDPKNPGGYTL 141  
DB 187 LAKQIGKEINLLGNVAYQQTENGKVCY-FSELIRLVAEIDGIERIRFTTSNPLEFT- 244  
QY 142 SDAYRYFALAREMKGNSTFPKPKVSNISFIHTWMN- - - - -DGLSIIISVIVNCVDFSLN 197  
DB 245 DDIIIEVYKDTPLKVSFLHLPVQSGSNKILNMKSYTTEDVTSIIKKLTARPDQIS- - 302  
QY 198 SISPYICGGAGVDAIEP- - - - -FDVLHKKPAYQSKGI- - - - -230  
DB 303 - - - - -SDFIVGFPGESEIDFEKTIETKINIFD-MSFSFIYSARPPTPASNMDDLDLKEKR 359  
QY 231 -AYSIPSNISLAFSLYHKVMGNQKRLNVQHVAE 264  
DB 360 RLYILOERINIQTLWMSRKMPGS-IQSVLVEGVSD 393

RESULT 18  
PURL\_METJA  
ID PURL\_METJA  
AC Q57657;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Probable amidophosphoribosyltransferase (EC 2.4.2.14) (Glutamine

DE phosphoribosylpyrophosphate amidotransferase) (ATASE) (GPATase).  
GN PURF OR MJ0204.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=8688087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodex A.,  
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.,  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
jannaschii.";  
RL Science 273:1058-1073(1996).  
CC -!- CATALYTIC ACTIVITY: 5-phospho-beta-D-ribose + diphosphate +  
CC L-glutamate = L-glutamine + 5-phospho-alpha-D-ribose 1-diphosphate  
CC + H(2)O.  
CC -!- PATHWAY: De novo purine biosynthesis; first step.  
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE  
CC PURINE/PRIMIDINE PHOSPHORIBOSYLTRANSFERASE FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 TYPE-2 GLUTAMINE AMIDOTRANSFERASE DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
CC EMBL; U67476; AB98188.1; -  
CC HSSP; P00497; IGPH.  
CC MEROPS; C44.UNW; -  
CC TIGR; MJ0204; -  
CC InterPro; IPR000583; GATase 2.  
CC InterPro; IPR000836; PRtransferase.  
CC InterPro; IPR002375; Pr/pyr.transf.  
CC Pfam; PF00156; Pribosyltran; 1.  
CC Pfam; PF00310; GATase 2; 1.  
CC TIGRPFAMs; TIGR01134; purF; 1.  
CC PROSITE; PS00103; PUR\_PYR\_PR\_TRANSFER; 1.  
CC PROSITE; PS00443; GATASE\_TYPE\_II; 1.  
CC Purine biosynthesis; Transferase; Glycosyltransferase;  
KW Glutamine amidotransferase; Complete proteome.  
FT INIT MET 0 BY SIMILARITY.  
FT ACT SITE 1 1 GATASE (BY SIMILARITY).  
SQ SEQUENCE 470 AA; 52056 MW; C8233EE969B08928 CRC64;  
  
Query Match 6.1%; Score 91.5; DB 1; Length 470;  
Best Local Similarity 20.6%; Pred. NO. 9.5;  
Matches 65; Conservative 47; Mismatches 111; Indels 93; Gaps 14;  
  
QY 51 VPSNFSVKETNVTIKNLALKKQVDSDIETKTDASVGINSPNFTIPY-----TAVQD 103  
DB 3 IFGIYSERLNVAKKIYYGL-----FALQHRGQEGAGIATSDGKNIHYKNIGLVTDVFN 58  
QY 104 NSV-NFNQGTIG-----YTFAEQTRVE-----IEGSVEEFDVKNPGGYTLSDAVRYPALA 151  
DB 59 ETQLNLFYIGIHGVRYSYTTGGKAVENQPFVKSFGNIAIHNGDLVNSDELR---RE 115  
QY 152 REMKGNSTPPK-----EKVNSIFHTV-----MRNDGLSLIIS 183  
DB 116 LEMKGIHFTSSDSEVIAQLLVRELLKTSKDIEAKNTLKKLVGAYSLIMFND--SLIA 173  
QY 184 V-----IVNVCVDSLNLNLSIPICGGAGVDA-----IEFDVLHIK-----FAYQSKLG 229

Db 174 VRDPWGFKPLCIGRDESNIIYISSDCALTTILDABFVKDIEPGEIIEIKDGIISHKLDYG 233  
QY 230 IAYSLPSNIS-----LFASLYY-----HKVNGNOFKNLNVQHVAEIA 266  
Db 234 VSEYNPVNDVPCIRGAATCMFEYVYFARPDSTIDGIVYKVRKRIKILAKEHPVDAD 293  
QY 267 SIPKITSAVATNLIGY 282  
Db 294 VVSPIDPSGVTFALGF 309  
  
RESULT 19  
Y075\_MYCGE  
ID Y075\_MYCGE STANDARD; PRT; 1024 AA.  
AC P47321; Q49190;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein MG075.  
GN MG075.  
OS Mycoplasma genitalium.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2097;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE=96026346; PubMed=7569993;  
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,  
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,  
RA Nguyen D.F., Dougherty T.R., Saudek D.M., Phillips C.A., Merrick J.M.,  
RA Tomb J.-F., Dougherty B.A., Bitt K.F., Hu P.-C., Lucier T.S.,  
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
RT "The minimal gene complement of Mycoplasma genitalium.";  
RL Science 270:397-403(1995).  
RN (2)  
RP SEQUENCE OF 1-156; 269-402; 643-736 AND 808-947 FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE=94075230; PubMed=8253680;  
RA Peterson S.N., Hu P.-C., Bitt K.F., Hutchison C.A. III;  
RT "A survey of the Mycoplasma genitalium genome by using random  
RT sequencing.";  
RL J. Bacteriol. 175:7918-7930(1993).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U39688; AAC71293.1; -  
CC EMBL; U01715; AAC43189.1; ALT\_INIT.  
CC EMBL; U02251; AAD12514.1; -  
CC EMBL; U01749; AAD10562.1; -  
CC EMBL; U01775; AAD10595.1; -  
CC TIGR; MG075; -  
KW Hypothetical protein; Transmembrane; Complete proteome.  
FT TRANSMEM 11 31 POTENTIAL.  
FT TRANSMEM 951 971  
SQ SEQUENCE 1024 AA; 116424 MW; 5B1540E6855CB554 CRC64;  
  
Query Match 6.1%; Score 91; DB 1; Length 1024;  
Best Local Similarity 22.7%; Pred. No. 27;  
Matches 62; Conservative 52; Mismatches 83; Indels 76; Gaps 18;  
  
QY 19 SLPNISSKA---INN- AKKYYGLYISGYKPSVSVFNSFKETNVTIKN-----LIAL 70  
DB 713 SLESVDSDFIKINGSFTLKYHG-----DNLNLLPNVH-----SLITKNVGVQIVNV 760  
QY 71 KQVDSD-IEYKTDASVGINP-----SNFTIPYTAVFQDMSVNFNGTIGTFFA 117

Db	761	NFHIDARLLTABLQNTVFSNPKVITKSVPVLSKSLFEV-WKTIF-ENSVNQILIKS3YTFK	818
Qy	118	EGTR--VEIEGSYE-EFDVKNPG---GYTSLDAYRYFALAREMKGNSTFPKEKYSN--	167
Db	819	DNLKFPFPKADGSSRLFEDLSKPDORVIPFAFVQYQ-FQLKKELIIPKETKEANSSPV	877
Qy	168	-SIFHTVMRNDG-----LSIISVIVNCVDFSLNNLSISPIICGAGV	209
Db	878	LKLYDAVKENDRQYRPNHHHDDLNRNPSLKQLEILLNLGDLKANNDFFDDTV-----V	932
Qy	210	DAIEPFDVLHIIFAYOSKLGII-AYSLSPNISLIF	241
Db	933	NALQY-----KTSFKSLTKVNSLIGIPINLFFF	959

```

RESULT 20
AIDA ECOLI
ID ID -AIDA_ECOLI STANDARD; PRT; 1286 AA.
AC Q03155;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Adhesin aidA-I precursor.
GN AIDA-I.
OS Escherichia coli.
OG Plasmid pIB6.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OC Escherichia
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 50-56.
RC STRAIN=O126:H27 / 2787;
RX MEDLINE=92326638; PubMed=1625582;
RA Benz I., Schmidt M.A.;
RT "AIDA-I, the adhesin involved in diffuse adherence of the
RT diarrhoeagenic Escherichia coli strain 2787 (O126:H27)", is
RT synthesized via a precursor molecule.";
RL Mol. Microbiol. 6:1539-1546(1992).
CC -1- FUNCTION: THIS IS AN ADHESION PROTEIN NECESSARY FOR THE DIFFUSE
CC ADHERENCE OF DIARRHEA-CAUSING ENTEROPATHOGENIC ESCHERICHIA COLI
CC TO EPITHELIAL CELLS.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; G65022; CAA46156.1; -.
CC DR PIR; S28634; S28634.
CC DR InterPro; IPR004899; Pertactin_sup.
CC DR Pfam; PF03212; Pertactin; 1.
CC KW Cell adhesion; Signal; Outer membrane; Plasmid.
CC FT SIGNAL 1 49
CC FT CHAIN 50 ? ADHESIN AIDA-I.
CC FT PROPEP 50 ? 1286
CC SQ SEQUENCE 1286 AA; 1332271 MW; B2A00F72AC05FB34 CRC64;

```

```

Qy 134 KNFGGYTSLDAYRYFALAREMKGNSFTPEKV--SNIFHTVMRNDGLSIISVIVWVCYDF 192
      :|::||::|:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Db    872 ---GDNSLTDRL-----VVKGNTSGSDIYYVNEDGGGQTDRDGINIIISVEGNSDAEF 921
      :|::||::|:||||:||||:||||:||||:||||:||||:||||:||||:|

Qy 193 SLNNLSISPICGA-----GVDAIEFFDLHIKFA----YQSCLKGIAYSILPSN 237
      |||:|||::||::||::||::||::||::||::||::||::||::||::||:|
Db 922 SLKN-----RVAGAYDYTLQKNESGDTNKGWYLTSHTPTSTOTRPENG---SYATN 973
      :|::||::|:||||:||||:||||:||||:||||:||||:||||:||||:|

Qy 238 ISLFASLYHKV-MGNQFNKLNVQHVAELASI-PKITSAVT-LN-----IGY 282
      |||:|||::||::||::||::||::||::||::||::||::||::||::||:|
Db 974 MALANSFLMDLNERNKQFRAMSDNTQPESASVWMKITGGISSCKLDNGQNKTINQFINQ 1033
      |||:|||::||::||::||::||::||::||::||::||::||::||::||:|

Qy 283 FGGEI 287
      ||::||:|
Db 1034 LGGDI 1038
      |||::||:|

RESULT 21
MSPI_PLAF3
ID MSPI_PLAF3 STANDARD; PRT; 1682 AA.
AC P19598; Q25921;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Merzoite surface protein 1 precursor (Merzoite surface antigens)
DE (PMMSA) (P190).
DE MSP-1.
GN Plasmodium falciparum (isolate ro-33 / Ghana).
OS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmidium.
OC NCBI_TaxId=5834;
OX [1]
RN SEQUENCE OF 1-1061 FROM N.A.
RX MEDLINE=88166657; PubMed=3327688;
RA Certa U., Rotmann D., Matile H., Reber-Liske R.;
RT "A naturally occurring gene encoding the major surface antigen
RT precursor p190 of Plasmodium falciparum lacks tripeptide repeats.";
RL EMBO J. 6:4137-4142(1987).
[2]
RN SEQUENCE OF 1032-1682 FROM N.A.
RX MEDLINE=95354793; PubMed=7628566;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of
RT merozoite surface antigen-1."
RL Exp. Parasitol. 81:47-54(1995).
CC CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC KDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to licens@isb-sib.ch).
CC -----
EMBL; M35727; AAA29715.1; -.
EMBL; Y00087; CAA68280.1; -.
EMBL; Z35326; CAA84555.1; -.
PIR; S06286; S06286.
DR InterPro: IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merzoite; Polyporein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1682 MEROZOITE SURFACE PROTEIN 1.
FT TRANSHEM 1666 1682 MEMBRANE ANCHOR (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 462 462 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 599 599 N-LINKED (GLCNAC... ) (POTENTIAL).

```



```
FT CARBOHYD 785 785 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 881 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 901 901 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 947 947 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1178 1178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1569 1569 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1682 AA; 192462 MW; C82A1E159948CAD6 CRC64;

Query Match 6.0%; Score 90.5; DB 1; Length 1682;
Best Local Similarity 21.0%; Pred. No. 54;
Matches 66; Conservative 42; Mismatches 110; Indels 97; Gaps 14;

Qy 4 KLFPTIINTVLVCLLSLNPNISSKAINNNAKYYGLYISGQYKPSVSVFSNFVKETNVI 63
Db 626 KIELFLKKAQLKDSIHVENIYKQ---NKPEPY----- 656
Qy 64 TKNLIALKDDVDSIE-----TKDASVCISNPSNFTIPYTAVFQDQNSVFNFTIGY 114
Db 657 ---LVLKKEVDKLFKEPTPKVKMDLKEQAV-----LSSITQPLVAASSETTEDGGHST--H 707
Qy 115 TFAEGRVIEGSEYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKK--KVSNSIPIHT 172
Db 708 TLSQSGETEVEETEE-----TVGHTTTVTI-----TLPPKVKVENSIEH- 749
Qy 173 VWRNDGLSIIIVNVVCDYFSLNNLSISPYIC-----GGAGVD-----AI 212
Db 750 -KSNDSQALTKTV---YLKDLDEPLTKSYICHKYLVSNSMDQKLEVLNLTPEENEL 805
Qy 213 EFDVHLHKFAYQSKGLTAYSLPSNIS-----LPASLYHKYKMGNOFKNLNVQHVAEIAS 267
Db 806 KSCORLDLFTIQQNNIPAMYSLYDSMNNDLQHLFFELYQKEMYYLHLKRENIHKLE 865
Qy 268 IPK-ITSAVATLINIG 281
Db 866 EPKQITGTSSTSPG 880

RESULT 22
DAAA_RICCN
ID_DAAA_RICCN STANDARD; PRT; 463 AA.
AC Q92H56.
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chromosomal replication initiator protein dnaA.
GN DAAA OR RC0916.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Maish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
CC -!- FUNCTION: Plays an important role in the initiation and regulation
CC of chromosomal replication. Binds to the origin of replication; it
CC binds specifically double-stranded DNA at a 9 bp consensus (dnaa
CC box): 5'-TTATC(C/A)(C/A)-3'. DnaA binds to ATP and to acidic
CC phospholipids (By similarity).
CC -!- SIMILARITY: BELONGS TO THE DAAA FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M90440; AAA21918.1; -
InterPro; IPR001343; Hemlysen_Ca_bind.
InterPro; IPR003355; RTXtoxin_N.
InterPro; IPR003995; RTXa.
```

```
CC or send an email to license@isb-sib.ch).
-----
EMBL; AE008645; AAL03454.1; -
DR InterPro; IPR001957; Bac DnaA.
DR Pfam; PF00308; bac_dnaA; 1.
DR TIGRFAMs; TIGR00362; DnaA; 1.
DR PROSITE; PS01008; DnaA; 1.
KW DNA replication; DNA-binding; ATP-binding; Complete proteome.
SQ NP_BIND 168 175 ATP (POTENTIAL).
SQ SEQUENCE 463 AA; 52942 MW; B48D2FE43A70A8EA CRC64;

Query Match 6.0%; Score 90; DB 1; Length 463;
Best Local Similarity 24.0%; Pred. No. 12;
Matches 42; Conservative 30; Mismatches 71; Indels 32; Gaps 10;

Qy 3 NKLKF--TIINTVLVCLLSLNPNISSKAINNNAKYYGLYISGQYKPSVSVFSNF--SVK 58
Db 40 SKVNFISSESLNTVLC-----APTDFVRDWIKSYSWI-----LQLFQHNNTIK 85
Qy 59 ETNVITKNLIALKDDVDSIEFTKTDASVIGS--NPSNF--TIPYTAVFQDQNSVFNFTIGY 114
Db 86 SIEIITKELPGTQTVTTELPTKTFADIGSSSELSENIFSTLDRFTDFNVFVGAPNELAY 145
Qy 115 TFAEGRVIEGSEYEEFDVKNP-----GGYTLSDAYRYFALAREMKGNSFTPKK 155
Db 146 AARAV--AESGAVSE---SNPLFYGGVGLGKTHLMHAIGWYIKOHN--ESRKV 194

RESULT 23
HLXA_ACTSU
ID_HLXA_ACTSU STANDARD; PRT; 956 AA.
AC Q00951;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hemolysin (Cytolysin II) (CLY-IIA) (HLY-IIA) (CYTC) (APPA).
GN APPA OR CLYIIA OR HLYIIA OR CYTC.
OS Actinobacillus suis.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Actinobacillus.
OX NCBI_TaxID=716;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3714;
RX MEDLINE=92267623; PubMed=1587585;
RA Burrows L.L., Lo R.Y.;
RT "Molecular characterization of an RTX toxin determinant from
RT Actinobacillus suis.";
RL Infect. Immun. 60:2166-2173(1992).
CC -!- FUNCTION: ONE OF THE VIRULENCE FACTORS OF A.SUIS. MIGHT BE A
CC SECRETED CYTOTOXIN, POSSIBLY THE EXTRACELLULAR HEMOLYSIN.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING
CC CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC
CC ACTIVITY.
CC -!- DOMAIN: THE THREE TRANSMEMBRANE DOMAINS ARE BELIEVED TO BE
CC INVOLVED IN PORE FORMATION BY THE CYTOTOXIN (BY SIMILARITY).
CC -!- PTM: PALMITOYLATED BY LKTC. THE TOXIN ONLY BECOMES ACTIVE WHEN
CC MODIFIED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M90440; AAA21918.1; -
InterPro; IPR001343; Hemlysen_Ca_bind.
InterPro; IPR003355; RTXtoxin_N.
InterPro; IPR003995; RTXa.
```

RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,  
RA Peterson J., Kervilange A.R., Quackenbush J., Salzberg S., Hanson M.,  
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,  
RA Uterback T., Watthey L., McDonald L., Artiach P., Bowman C.,  
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,  
RA Smith H.O., Venter J.C.;  
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia  
RT burgdorferi.";  
RL Nature 390:580-586(1997) .  
RN [3].  
RP SEQUENCE OF 1-28 FROM N.A.  
RC STRAIN=212;  
RX MEDLINE=93366157; PubMed=8359672;  
RA Old I.G., Margarita D., Saint-Gironis I.;  
RT "Unique genetic arrangement in the dnaA region of the Borrelia  
RT burgdorferi linear chromosome: nucleotide sequence of the dnaA  
RT gene.";  
RL FEMS Microbiol. Lett. 111:109-114(1993).  
CC -1- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME  
CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.  
CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.  
CC THE BETA CHAIN IS REQUIRED FOR INITIATION OF REPLICATION ONCE IT  
CC IS CLAMPED ONTO DNA, IT SLIDES FREELY (BIDIRECTIONAL AND ATP-  
CC INDEPENDENT) ALONG DUPLEX DNA (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
CC + [DNA] (N).  
CC -1- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND  
CC THETA CHAINS) THAT CAN REPAIR SHORT GAPS CREATED BY NUCLEASE  
CC IN DUPLEX DNA. FOR EFFICIENT REPLICATION OF THE LONG, SINGLE-  
CC STRANDED TEMPLATES, POL III REQUIRES THE AUXILIARY CHAINS BETA,  
CC GAMMA, AND DELTA (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U04527; AAA58942.1; -.  
DR EMBL; AE001149; AAB91514.1; -.  
DR PIR; S34948; S34948.  
DR TIGR; BB0438; -.  
DR InterPro; IPR001001; DNA\_polIII\_beta.  
DR Pfam; PF00712; DNA\_pol3\_beta\_1.  
DR Pfam; PF02767; DNA\_pol3\_beta\_2.  
DR Pfam; PF02768; DNA\_pol3\_beta\_3.  
DR SMART; SM00480; POL3BC; 1.  
DR TIGRFAMs; TIGR00663; dnan; 1.  
KW Transferase; DNA-directed DNA polymerase; DNA replication;  
KW Complete proteome.  
KW CONFLICT 86  
FT SEQUENCE 385 AA; 44641 MW; 87CE040D369BD0A5 CRC64;  
SQ  
Query Match 6.0%; Score 89.5; DB 1; Length 385;  
Best Local Similarity 21.5%; Pred.No. 10;  
Matches 69; Conservative 53; Mismatches 113; Indels 87; Gaps 19;  
Qy 11 NTVLVCLLSLPNTSSSKAINNNKYYGLYISGQKPSVSVFNSFYKTNVITKNL--- 67  
Db 4 NTFPIC-----ETNQINWEIEKAGIILNRMNDIWSALL-IEVKKSLNIKSTDRN 54  
Qy 68 IALKKQVDSITETKIDASVGISNPSNFTPIPTAVFDQNSVNPNGTIGTYTFAE-GTRVIEIG 126  
Db 55 IFFESTI-SIVSETDFKV-LINASF---YDAV---KAFNYFKIKIVFNENNSKLEIMG 106  
Qy 127 SY---EEFD-VKNPG-GYTLSDAVRY-----FALAREMKGNF-----T 160  
Db 107 ELNDEKEIEDHUKPTFSVEEIEYNYDMWNEDYTFGIEIKQKSFKKVINRTAFSAHLD 166  
Qy 161 PKEKVSNSIFHTVMRNDGLSIIS-----VIV--NVCYDFSLNLSISPYTCGG 206

```
DR FRAM; PF02742; DNA_PoL3_beta_1;
DR FRAM; PF02767; DNA_PoL3_beta_2;
DR FRAM; PF02788; DNA_PoL3_beta_3;
DR SMART; SMO0480; POL3BG;
DR TIGRFAMs; TIGR00663; dnan;
DR TransFamS; Dna-directed DNA polymerase; DNA replication; Complete proteome.
KW TRANSFERASE
KW CONFLICT
PT CONFLICT
SQ SEQUENCE      86          N -> S (IN REF. 2).
                   385 AA;   44641 MW;   87CE040D369BD0A5 CRC64;

Query Match              6.0%; Score 89.5; DB 1; Length 385;
Best Local Similarity    21.5%; Pred.No.10;
Matches                  69; Conservative 52; Mismatches 113; Indels 87; Gaps 19;
```

QY                  11 NTVLVCLLSPLNISSKAIINNKKYYGLVIISGYKPFSVSFNFSVKETNVITKL-- -- 67

Db                 4 NTFIC-----ETNQIWEIEKAAGILNRNMNDIWSALL-IEVKSNLIIKSTDRN 54

QY                68 IALKKOVDSIETKTIDASGVISNPNSFTIPTAYVFQDNSVMNFNGTIGTYFAE-GTRVEIEG 126

Db                55 IPFEISTI-SIVSETDFKV-LINASNF---YDAV---KAENFYKIKIVFNENNNSKLEIMG 106

QY                127 SY----EEFD--WKDPG-GYTLDAYRY-----PALAREMKGNSF-----T 160

Db                107 ELNDEKESEEDHLHKPTFSYEIEINYNYDMWNEDYTGTEIIOKSKFKVINRIAFSAHLD 166

QY                161 PKEKVSNISIFHTWVRNDGLSIIS-----VTV--NCVYDFSNNLSISPYCIGG 206

```

Db 167 ESKVNLNGVYFSDKEDSKLLVSTNGHRMSCKTEVIVEEDVNFVTPVKIFNFKLHLSMG 226
Qy 207 AGVDAIEFFD-VLHKFAYQSKGLIAYSLPNSISLFSALY-----YHKVNGNOFKNLNV 259
Db 227 EGMVKIKESDKFYVEF-----DNYKIACSLINGNYPDYKSIIPKQKN----- 270
Qy 260 QHVAELASIPKITSATVATLNI 280
Db 271 ---KSLVSLGILKORLARVNL 288

```

## RESULT 25

```

ID CDGT_BACS2 STANDARD; PRT; 703 AA.
AC P31746;
DT 01-JUL-1993 (Rel. 26, Created)
*DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Cyclomaltodextrin glucanotransferase precursor (EC 2.4.1.19)
DE (Cyclodextrin-glycosyltransferase) (CGTase).
GN CGT.
OS Bacillus sp. (strain 1-1).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=29334;
[1]
SEQUENCE FROM N.A., AND SEQUENCE OF 30-52.
RP Schmid G., Englbrecht A., Schmid D.;
RT "Cloning and nucleotide sequence of a cyclodextrin glycosyltransferase
RL gene from the alkalophilic Bacillus 1-1.";
RL (in) Huber O., Szejtli J. (eds.);
RL Proceedings of the fourth international symposium on cyclodextrins,
RL pp.71-76, Kluwer Academic Publishers, Dordrecht and Boston (1988).
CC -1- CATALYTIC ACTIVITY: Degrades starch to cyclodextrins by formation
CC of a 1,4-alpha-D-glucosidic bond.
CC -1- COPACTOR: BINDS TWO CALCIUM IONS.
CC -1- SUBUNIT: MONOMER.
CC -1- MISCELLANEOUS: CGTASE MAY CLEAVE TWO PROTEIN DOMAINS: THE ONE
CC IN THE AMINO-TERMINAL SIDE CLEAVES THE ALPHA-1,4-GLUCOSIDIC BOND
CC IN STARCH, AND THE OTHER IN THE C-TERMINAL SIDE CATALYZES OTHER
CC ACTIVITIES, INCLUDING THE RECONSTITUTION OF AN
CC ALPHA-1,4-GLUCOSIDIC LINKAGE FOR CYCLIZING THE
CC MALTOOLIGOSACCHARIDE PRODUCED.
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
DR PIR; S26399; ALBSX1.
DR HSSP; P31797; 1CYG.
DR InterPro; IPR000461; Alpha_amylase.
DR InterPro; IPR002044; CBD_4.
DR InterPro; IPR002909; IPT_TIG.
DR Pfam; PF00128; alpha-amylase; 1.
DR Pfam; PF00686; CBM_20; 1.
DR Pfam; PF01833; TIG; 1.
DR Pfam; PF02806; alpha-amylase_C; 1.
DR ProDom; PD001568; CBD_4; 1.
KW Transferrase; Glycosyltransferase; Calcium; Signal.
FT SIGNAL 1 29
FT CHAIN 30 703 CYCLOMALTODEXTRIN GLUCANOTRANSFERASE.
FT DOMAIN 30 160 A1.
FT DOMAIN 161 224 B.
FT DOMAIN 225 428 A2.
FT DOMAIN 429 516 C.
FT DOMAIN 517 600 D.
FT DOMAIN 601 703 E.
FT DISULFID 68 75 BY SIMILARITY.
FT ACT_SITE 251 251 BY SIMILARITY.
FT ACT_SITE 279 279 BY SIMILARITY.
FT ACT_SITE 350 350 BY SIMILARITY.
SQ SEQUENCE 703 AA; 78663 MW; 4D973B21D0D9B0A CRC64;

```

Query Match 6.0%; Score 89.5; DB 1; Length 703;  
 Best Local Similarity 19.7%; Pred. NO. 22;  
 Matches 62; Conservative 40; Mismatches 101; Indels 111; Gaps 13;

```

Qy 1 MNKKLKFTIINTVLVCLLSLPNTSSKAIN--NNAKKYGLYISGQY-----KPSVSV 51
Db 4 LNDELKTLISFIFLLLSLPTVAEADVTKVNSKDVIVQIVTDREFSDGNPGNPGSGAI 63
Qy 52 FSNFSVETNVTIKVIALKK-----DVDSTETKTDSVSGIS-----N 89
Db 64 FS-----QNCIDLHKYCGDWOGIIDKINDGLYLDLGTALTALWISQPVENVYALH 112
Qy 90 PSNFTI-----PYTAVFQDSNVFNCTIGTGTFAEGTRV-----E 123
Db 113 PSGYTSYHGWARDYKKTNPYGNFDD-----FDRLMSTAHSNGIKVIMDFTPNHSSPALE 168
Qy 124 IEGSYEBFDVKNPGGYTLSDAYRYFALAREMKNGSFTPKKVSNSIFHTVMRNDGLSIIS 183
Db 169 TNPYVENGAIYDNGALL-----GNYSNDQQ-----NLFHNGGTDFFSSYED 210
Qy 184 VIVNVCY---DFSNNLSISPYICGG-----AGVDAIEFFDVLHIFAYQSKGLIAYSL 234
Db 211 SIYRNLYDLADYLDLNTVMDQYLKESIKFWLDKIGIDIRVDVAVKHMSEGWQT----- 262
Qy 235 PSNISLFSALYYHK 248
Db 263 ----SLMSEIYSHK 272

```

## RESULT 26

```

ID -FRET3 YEAST STANDARD; PRT; 711 AA.
AC Q08905;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ferric reductase transmembrane component 3 precursor (EC 1.6.99.13)
DE (Ferric-chelate reductase 3).
GN FRET3 OR YOR381W OR O6754.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX NCBI_TaxID=4932;
[1]
SEQUENCE FROM N.A.
RA Delius H., Hebling U., Hofmann B.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: NADH + 2 Fe(3+) = NAD(+) + 2 Fe(2+).
CC -1- COPACTOR: FAD (PROBABLE).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE FRE / CYBB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; Z75289; CAA99713.1; -.
DR SGB; S0005908; FRET3.
DR InterPro; IPR002916; Ferric_reduct.
DR Pfam; PF01794; Ferric_reduct; 1.
DR KX Oxidoreductase; Electron transport; Transmembrane; Iron transport;
KW FAD; NAD; Glycoprotein; Signal; Multigene family.
FT SIGNAL 1 20
FT CHAIN 21 711 FERRIC REDUCTASE TRANSMEMBRANE COMPONENT
FT NP_BIND 479 485 FAD (POTENTIAL).
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 167 187 POTENTIAL.
FT TRANSMEM 238 258 POTENTIAL.
FT TRANSMEM 281 301 POTENTIAL.
FT TRANSMEM 354 374 POTENTIAL.
FT TRANSMEM 377 397 POTENTIAL.

```

```

FT TRANSMEM 399 419 POTENTIAL.
FT TRANSMEM 464 484 POTENTIAL.
FT TRANSMEM 540 560 POTENTIAL.
FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 108 108 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 614 614 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 669 669 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 711 AA; 80589 MW; 362E18F7A34C165 CRC64;

Query Match
Best Local Similarity 18.1%; Score 89.5; DB 1; Length 711;
Matches 79; Conservative 58; Mismatches 130; Indels 169; Gaps 20;

QY 9 INTVLVCLLSLPNTSSKAINNNAKKYK-----LYISG 43
D 6 LCGSILLCLCSGASAPAKT-----KMYGKLPLVLTACMGVLGEVTWEYSDDLISPP 59
QY 44 --QYKPSVS-----VFSNFSVK--ETNVITKNLIALKKD-----VDSIETKTDSVGINP 90
D 60 ACTYEPALQSLMCLYCSLNEKGYSNRTFEKTAFAIKEDCAYYTONLQNMNAD--FYNM 117
QY 91 SNFTTPTAVFODNSVNFNGTI--GYTFAEGRTRVEIEGSEYEFVKNP--GGYTLSDAYRY 147
D 118 LNNGTYYIIQYSEGSANLTYPIEMDAQVRENYYYSYHGFYANDIGHTYGGIICAYFVG 177
QY 148 PALAREMGKNSPT-----KEKVSNSI--FHTV-----MRND 177
D 178 MILASILHVSITPTKALFKQLRVYVRRYLTPTINGKSHASSPSYLYKIFGFLPTRSE 237
QY 178 GLSILSVIV-----NVCYD-----PSLNNLSISPICGAGVDATFEFVLIHKA- 223
D 238 GVILGLYLVLHVTFLAYGYQYDPYNLI FDSRREQIARYVADRSGVLAFAPFLIAL-FAG 296
QY 224 -----YQSKL-----GLAYSLLPS-----236
D 297 RNNFLFISGVKYTFIMPHKWLGRMFLDAVHGAATSYISVFKDWAASKEEYTWQFG 356
QY 237 -----NTSLFASLYV-----HKVMGNQFNVLNVQHVAAELASIPKITSAVA- 276
D 357 VAALCIVGVVVFSLAMERKFFYEAFPLHLHVLGALFFYTCEWHVVELSGIEWIAAIAI 416
QY 277 -----TLNIGYFG 284
D 417 WTIDRLIRIVRSYFG 432

RESULT 27
PECL_PIG
ID PECL_PIG STANDARD; PRT; 740 AA.
AC Q95242;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Platelet endothelial cell adhesion molecule precursor (PECAM-1)
DE (CD31 antigen).
GN PECAM1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxId=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Nasu K.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE EXPRESSED ON
CC PLATELETS AND AT ENDOTHELIAL CELL INTERCELLULAR JUNCTIONS.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -!- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X98505; CAA67129.1; -.
DR InterPro; IPR003599; IG.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003600; IG_Like.
DR Pfam; PF00047; IG; 4.
DR SMART; SM00409; IG; 2.
DR SMART; SM00410; IG_Like; 1.
KW Immunoglobulin domain; Glycoprotein; Cell adhesion; Transmembrane;
KW Repeat; Signal.
FT SIGNAL 1 27
FT CHAIN 28 740
FT DOMAIN 28 602
FT TRANSMEM 603 621
FT DOMAIN 622 740
FT DOMAIN 50 116
FT DOMAIN 145 213
FT DOMAIN 249 311
FT DOMAIN 340 394
FT DOMAIN 425 484
FT DOMAIN 517 580
FT DISULFID 57 109
FT DISULFID 152 206
FT DISULFID 256 304
FT DISULFID 347 387
FT DISULFID 432 477
FT DISULFID 524 573
FT CARBOHYD 52 52
FT CARBOHYD 84 84
FT CARBOHYD 284 284
FT CARBOHYD 301 301
FT CARBOHYD 320 320
FT CARBOHYD 357 357
FT CARBOHYD 372 372
FT CARBOHYD 436 436
FT CARBOHYD 456 456
FT CARBOHYD 552 552
SQ SEQUENCE 740 AA; 82378 MW; F312DC62C4B4A217 CRC64;

Query Match
Best Local Similarity 19.2%; Score 89.5; DB 1; Length 740;
Matches 60; Conservative 51; Mismatches 93; Indels 109; Gaps 12;

QY 22 NISSKAINNNAKKYK-----GLYISGQYKPSV-----49
D 84 NVSSTK-----NTSEYFISEARVNSGRYKCTVILNKEKTTAEYKVVVEGVSNPRTVLDK 139
QY 50 -SVFSNFSVKETNVTK-----NLIALKQVDSITETKTDSVGSISNPSNFTIPTAVFQD 103
D 140 KEVIEGGVVKVTCSPBEKPPVHFIEKFLNVDRVQRREKTNANNQVLETFVEEQD 199
QY 104 NSVNFNGTIGYTFAGTRVEIEGSEYEFVKNPGGYTLSDAYRYFALAREMKGNSFTPE 163
D 200 RVILFSCQANVIF--GTRVEISDS-----VRSDLVTVRE 231
QY 164 KVSNSIFH-----TVMRNDGLSIISIVNVICYDFSLNNLSISPICGAGVDATFEFFVL 218
D 232 SFSNPKFHISPKGVIEGDQL-LIKCTIQVTH-----QAQSFPEII 271
QY 219 HIKAYQSKLGIAYSLPSNISLFA-----SLYHKVMGNQFNKLN--VQHVAAELAS 268
D 272 ----IQDKEIVAHNRNGSEAVISVMATVEHNSNYTKVEASRISKVSSIMWNITELFSR 327
QY 269 PKITSATVNLIG 281
D 328 PKLKSSATRLDQG 340

```



RP SEQUENCE FROM N.A.  
RX MEDLINE=86136024; PubMed=3004972;  
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,  
RA Stunnenberg H., Bujard H.;  
RT "Polymorphism of the precursor for the major surface antigens of  
RT Plasmodium falciparum merozoites: studies at the genetic level.";  
RL EMOB J. 4:3823-3829(1983).  
RN [2]  
RP REVISIONS, SEQUENCE FROM N.A.  
RA Pan W., Tolle R., Bujard H.;  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
CC (Potential).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X03371; CAA27070.1; -.  
DR PIR; A25120; SAZQK1.  
DR InterPro; IPR000561; EGF-like.  
DR Pfam; PF00008; EGF; 1.  
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;  
KW Transmembrane; GPI-anchor.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 1630 MEROZOITE SURFACE PROTEIN 1.  
FT DOMAIN 67 84 TRIPEPTIDE SG(tp) REPEAT.  
FT TRANSMEM 1614 1630 MEMBRANE ANCHOR.  
FT CARBOHYD 97 97 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 755 755 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 759 759 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 774 774 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 835 835 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 911 911 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 955 955 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1049 1049 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1156 1156 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1436 1436 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1517 1517 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 1630 AA; 187289 MW; ADBDEC3CE0A46322 CRC64;  
  
Query Match 6.0%; Score 89.5; DB 1; Length 1630;  
Best Local Similarity 19.0%; Pred. No. 62;  
Matches 55; Conservative 52; Mismatches 127; Indels 55; Gaps 10;  
  
QY 20 LPNISSKAINNAKYYGLISQYKPSVSVFNSFKVETNVTIKNLIAKQDVSTET 79  
DB 619 IDELKKTQILKNVELKHNHVPNSYK-----QENKQBPYYLVLKKEIDKLKV 667  
QY 80 KTDASVGSINPSNFTPIYTAVFQDMSVNPNGTIGYTFAGTVELEGSEVEEDVKNPGGY 139  
DB 668 PMPKVESLINEKNIK-TEGOSDMS-----EPSTGEITGQATKPGQOAGSA 715  
QY 140 TLDAYRYFALAR-----SINLISIPYICGGAGVDAIEFFDLVHLTKFAYQSLGATYSLPNI 187  
DB 716 LEGDSVQAQKQKQAPVPVPVPEAKAQVTPPAPVNNKT-ENVSKLDYLEKLYEFLN 774  
QY 188 ---VCYDF-----SINLISIPYICGGAGVDAIEFFDLVHLTKFAYQSLGATYSLPNI 238  
DB 775 TSYICHKYLIVSHSTNKEIKQYKITKEESKLSGCDPLDLFNQNNIPVWYSMFDSL 834  
QY 239 S-----LFASLYYHKVMGNQFK-----NLNVQHVAAELASIPKITSAVATLN 279

DB 835 NNSLSQLFMEIYEKEMVCMVNLKLDNDKIKNLLBEAK--KVSTSVKTLIS 881  
  
RESULT 30  
MSPI\_PLAFW  
ID MSPI\_PLAFW STANDARD; PRT; 1639 AA.  
AC P04933,  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)  
DE (PMSA) (P195).  
GN MSP-1.  
OS Plasmodium falciparum (isolate Wellcome).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5948;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86014355; PubMed=2995820;  
RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,  
RA Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,  
RA Freeman R.R.;  
RT "Primary structure of the precursor to the three major surface  
RT antigens of Plasmodium falciparum merozoites.";  
RL Nature 317:270-273(1985).  
RN [2]  
RP REVISIONS.  
RA Holder A.A.;  
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
CC (Potential).  
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42  
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X02919; CAA26676.1; -.  
DR PIR; A24594; A24594.  
DR InterPro; IPR000561; EGF-like.  
DR Pfam; PF00008; EGF; 1.  
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;  
KW Transmembrane; GPI-anchor.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 1639 MEROZOITE SURFACE PROTEIN 1.  
FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 268 268 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 764 764 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 768 768 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 783 783 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 844 844 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 920 920 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 964 964 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1058 1058 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1445 1445 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1526 1526 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 1639 AA; 187618 MW; 2C255B6616C87F66 CRC64;  
  
Query Match 6.0%; Score 89.5; DB 1; Length 1639;  
Best Local Similarity 19.0%; Pred. No. 63;  
Matches 55; Conservative 52; Mismatches 127; Indels 55; Gaps 10;  
  
QY 20 LPNISSKAINNAKYYGLISQYKPSVSVFNSFKVETNVTIKNLIAKQDVSTET 79  
DB 628 IDELKKTQILKNVELKHNHVPNSYK-----QENKQBPYYLVLKKEIDKLKV 676



```

Query Match      5.9%; Score 88.5; DB 1; Length 405;
Best Local Similarity 17.7%; Pred. No. 13;
Matches 68; Conservative 58; Mismatches 135; Indels 123; Gaps 16;

Qy      2 NNKLKFTIINT-----VLVCLLSLPNITSSKAINNAKKYGYIISGQ----- 44
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 34
PBBP_BUCAI
ID _BUCAI STANDARD; PRT; 760 AA.
AC P57296;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Penicillin-binding protein 1B (PBP-1b) (PBP1b) (Murein polymerase)
DE [Includes: Penicillin-insensitive transglycosylase (EC 2.4.2.-)
DE (Peptidoglycan TGase); Penicillin-sensitive transpeptidase
DE (EC 3.4.-) (DD-transpeptidase)].
DE MRCB OR BU200.
GN Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RT Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RA "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -!- FUNCTION: CELL WALL FORMATION, SYNTHESIS OF CROSS-LINKED
CC PEPTIDOGLYCAN FROM THE LIPID INTERMEDIATES. THE ENZYME HAS A
CC PENICILLIN-INSENSITIVE TRANSGLYCOSYLASE N-TERMINAL DOMAIN
CC (FORMATION OF LINEAR GLYCAN STRANDS) AND A PENICILLIN-SENSITIVE
CC TRANSPEPTIDASE C-TERMINAL DOMAIN (CROSS-LINKING OF THE PEPTIDE
CC SUBUNITS) (BY SIMILARITY).
CC -!- PATHWAY: FINAL STAGES IN PEPTIDOGLYCAN SYNTHESIS.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
CC TRANSGLYCOSYLASE FAMILY.
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE
CC TRANSPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on it
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercia
CC entities requires a license agreement (see http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP001118; BAB12917.1; -
DR InterPro; IPR001264; GT 51.

```



DR InterPro; IPR001460; Transpeptidase.  
DR Pfam; PF00905; Transpeptidase; 1.  
DR ProDom; PD001895; GT 51; 1.  
KW Peptidoglycan synthesis; Cell wall; Transferase; Glycosyltransferase;  
KW Hydrolase; Multifunctional enzyme; Transmembrane; Signal-anchor;  
KW Antibiotic resistance; Complete proteome.  
FT DOMAIN 1 8 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 9 29 SIGNAL-ANCHOR (TYPE II MEMBRANE PROTEIN)  
FT DOMAIN 30 760 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 136 308 TRANSGLYCOSYLASE.  
FT DOMAIN 392 684 TRANSPEPTIDASE.  
FT ACT SITE 451 451 ACYLATED BY PENICILLIN (BY SIMILARITY).  
SQ SEQUENCE 760 AA; 88002 MW; 00B2CSB51F6947D3 CRC64;  
  
Query Match 5.9%; Score 88.5; DB 1; Length 760;  
Best Local Similarity 24.4%; Pred. No. 29;  
Matches 78; Conservative 36; Mismatches 117; Indels 89; Gaps 17;  
  
Qy 2 NKKLFTTIINTVLVCLLSLPNISSS---ETNVITKVL---TALKKQVDS-----IETKTDAS 84  
Db 431 SSKPEFGYNRAKTRRSIGSLSPITVLTALSQPEKYHLNTWISNYPLSLIKLDSGQYWT 490  
  
Qy 47 PSVSFVSFVSFK-----ETNVITKVL---TALKKQVDS-----IETKTDAS 84  
Db 491 PKNMNF--FSKKVLLLDALIHISINPTVNLINIGLKKLVDSWLLGLGSKKYITPLPSIS 549  
  
Qy 85 VGINSPNFTIPTYTAVFO-----DQVFIIGGGYKSSLSVRSIISDDGKLVQNLQPSIHIESSEASY 607  
Db 550 LGAINLTPFEI--AQVFIIGGGYKSSLSVRSIISDDGKLVQNLQPSIHIESSEASY 607  
  
Qy 122 VEIEGSEYEEFDVQNPGGYTLSDAVRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGLSI 181  
Db 608 LTLGQMQQV--VKSGTAKSLGTIPKEFSLA---GKTGTNNLVN-----NFVGLDGKQI 657  
  
Qy 182 ISVIVNVCYDPSLNLSIPYICGAGVDATFEFDVLHIKFAYGSKLGIAYSLNSISLF 241  
Db 658 --VITWIGRD---NNHTRLYSSSGA-----MQIKRYLQYQRPVPLVLKAPNNIMF 705  
  
Qy 242 ASLYYHKVMGNQKLNLYQH 261  
Db 706 ----YINNIGELFCCKNNQH 721  
  
RESULT 35  
OP66 NEIGO STANDARD; PRT; 238 AA.  
AC Q05033;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Opacity protein OPA66 precursor (Fragment).  
OS Neisseria gonorrhoeae.  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=485;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VPI;  
RX MEDLINE=93178439; PubMed=8440254;  
RA Kupsch E.-M., Knepper B., Kuroki T., Heuer I., Meyer T.F.;  
RT "Variable opacity (Opa) outer membrane proteins account for the cell  
tropisms displayed by Neisseria gonorrhoeae for human leukocytes and  
epithelial cells";  
RL EMBL J. 12:641-650(1993).  
CC -!- FUNCTION: IMPLICATED IN A NUMBER OF ADHERENCE FUNCTIONS. OPA  
CC PROTEINS ARE IMPLICATED IN PATHOGENESIS AND ARE SUBJECT TO PHASE  
CC VARIATION.  
CC -!- SUBCELLULAR LOCATION: Outer membrane.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z18941; CAA79374.1; -.  
DR PIR; S28619; S28619.  
DR InterPro; IPR003394; Porin opacity.  
DR Pfam; PF02462; Opacity; 1.  
KW Outer membrane; Multigene family; Signal.  
FT NON TER 1 1 POTENTIAL.  
FT SIGNAL 2 >238 OPACITY PROTEIN OPA66.  
FT CHAIN 238 238  
FT NON TER 238 238  
SQ SEQUENCE 238 AA; 26888 MW; B6E38AF1585263AA CRC64;  
  
Query Match 5.9%; Score 88; DB 1; Length 238;  
Best Local Similarity 24.1%; Pred. No. 7.5;  
Matches 55; Conservative 26; Mismatches 79; Indels 68; Gaps 14;  
  
Qy 81 TDASVG-ISPNSFTIPTYTAVFDQNSVFNFTGTYTFAEGRVIEGSEYEEFDVKNPGGY 139  
Db 30 TDPKSGKLSVSD---YERNIRTHSIHPRVSGYDFG-GWR--IAADYARYKWNDSKY 82  
  
Qy 140 TLS--DAYRFALAREMKGNSFTPK-EKVSNSIFHTVMRNDGLSIISVIVNVCYDPSLNN 196  
Db 83 SVSINKLQR-----RTSGNRRDRKTKENQENGSPHAV--SSIGLSAV-----YDFKLDN 129  
  
Qy 197 LSIISPYI-----CCGAGVDAIEFFDVLHIKFAYSK-----227  
Db 130 -KFPYICARVAYGHVHRHSIDSTKKTFTFLTAGAGTDPVSSPYKNTQAHQESNIR 188  
  
Qy 228 -----LGAYSIPSNISLFASLYYH---KVMGNQKFNLMVQHVAE 265  
Db 189 RVGLGIAGVGFDTIPNLTLDAGYRHNWGRLENTREKT-----HEASL 232  
  
RESULT 36  
SD55 CAEEL STANDARD; PRT; 334 AA.  
AC Q09572;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Serpentine receptor class delta 55 (Srd-55 protein).  
GN SRD-55 OR K02A2.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Hallsworth K.;  
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP REVISIONS.  
RA Waterston R.;  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- SIMILARITY: BELONGS TO THE C.ELEGANS RECEPTOR-LIKE PROTEIN SRB  
CC FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U23171; AAC46702.2; -.  
DR WormPep; K02A2.2; CE21013.



```

01-JUN-1994 (Rel. 29, Created)
01-JUL-1995 (Rel. 32, Last sequence update)
15-JUL-1998 (Rel. 36, Last annotation update)
MYOSIN-3 isoform.
MYO3 OR YKL129C.
Saccharomyces cerevisiae (Baker's yeast).
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NCBI_TaxID=4932;
[1]
SEQUENCE FROM N.A.
STRAIN=CRY3;
MEDLINE=95246144; PubMed=7728870;
Goodson H.V., Spudich J.A.;
"Identification and molecular characterization of a yeast myosin I.";
Cell Motil. Cytoskeleton 30:73-84(1995).
[2]
SEQUENCE FROM N.A.
Rad M.R., Xu G., Kirchrath L., Fritz C., Keuchel H., Hollenberg C.P.;
Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
-1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
-1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL; S76960; AAB34124.1; -
CC EMBL; Z28129; CAA81970.1; -
CC PIR; S37958; S37958.
CC HSPSP; P08739; LMND.
CC SGD; S0001612; MYO3.
CC InterPro; IPR001452; SH3.
CC InterPro; IPR001609; myosin_head.
CC Pfam; PF00018; SH3; 1.
CC Pfam; PF00063; myosin_head; 1.
CC PRINTS; P000193; MYOSINHEAVY.
CC ProDom; PD000066; SH3; 1.
CC ProDom; PD000355; myosin_head; 1.
CC SMART; SM00242; MYSC; 1.
CC SMART; SM00326; SH3; 1.
CC PROSITE; PS50002; SH3; 1.
CC Myosin; Actin-binding; ATP-binding; SH3 domain.
CC DOMAIN 1 ? MYOSIN HEAD-LIKE.
CC FT DOMAIN 1121 1183 SH3.
CC FT NP BIND 130 137 ATP (POTENTIAL) .
CC FT DOMAIN 1112 1117 POLY-PRO.
CC FT DOMAIN 1260 1272 ASP-RICH (ACIDIC).
CC FT CONFLICT 95 96 RK -> G (IN REF. 2).
CC FT CONFLICT 169 169 T -> NP (IN REF. 2).
CC FT CONFLICT 263 263 T -> S (IN REF. 2).
CC FT CONFLICT 266 270 TIDDV -> OLMR (IN REF. 2).
CC FT CONFLICT 917 919 RLV -> VG (IN REF. 2).
CC FT CONFLICT 1022 1022 A -> R (IN REF. 2).
CC SEQUENCE 1273 AA; 142780 MW; 805D07E5B8D24233 CRC64;
SQ
Query Match 5.9%; Score 88; DB 1; Length 1273;
Best Local Similarity 19.0%; Pred. No. 59;
Matches 59; Conservative 50; Mismatches 120; Indels 82; Gaps 12;
QY 4 KUKFTIINTVLCLSLPNISSKAINNAKKYGLYI----SGQYKPSVSPFSVKE 59
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 160 KIKDMVLATL-----LESFGCAKTLRNNSRRHGKYLEIKFNSQPEPCAGNITNYLLEK 214
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 60 TWVITKGLALKQDVDSIETKTDASGVISNPNSFTIPYTAVDQNSVFNFTIGYTFAG 119
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 215 QRVVQ-----IKERNF-----HIFYQFTKG 236
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 120 TRVEIGSYEE-FDVKNPGGY-----TLSDAYRYFALAREMKGNSFTPKKYS 166
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```
Db 237 A-----SDTYKQMGVQMEQYIYTAAGCTTADTIDDVDKDYEGTLEARNRTGLVQEQ-- 290
Qy 167 NSIFHTVMRNDGLSIISVIVNVCYDFSLNLSISPIYCGAGVDAIEFFDVL---HIKFA 223
Db 291 DOIFRMLAAILWIGNISPIENEENAGVQDTSVDFVAYLLQVDASLLVKCLIVERIMQTS 350
Qy 224 YQSKLGIAYSLPN-----ISLFASLYHKVMGNQFNLVQHVAEHLASIPKITSVA 276
Db 351 HGMKRGSVYHVPNVPQATVRDALAKAIYNNLFDWIVDRVNV-----SLQAPPGADKSG 406
Qy 277 TLNIGYFGGEI 287
Db 407 ILDI--YGFEEI 415

RESULT 40
GTFB_STRMU
ID_GTFB_STRMU STANDARD; PRT: 1476 AA.
AC P08987; O69381; O69384; O69387; O69390; O69396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN GTFB.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]_TaxID=1309;
RP SEQUENCE FROM N.A.
RC STRAINS-GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4239, MT4245, MT4251, MT4467, and MT8148;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC OF THEIR ABILITY TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -!- CATALYTIC ACTIVITY: Sucrose + {(1,6)-alpha-D-glucosyl}(N) = D-
CC fructose + {(1,6)-alpha-D-glucosyl}(N+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DISEASE: DENTAL CARIES.
CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S.MUTANS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M17361; AAA88588.1; -
DR EMBL; D88651; BAA26101.1; -
DR EMBL; D88654; BAA26105.1; -
DR EMBL; D88657; BAA26109.1; -
```

```
DR EMBL; D88660; BAA26113.1; -
DR EMBL; D89977; BAA26119.1; -
DR PIR; B33135; B033135.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; GH 70.
DR Pfam; PF01473; CW binding_1; 13.
DR Pfam; PF02324; Glyco hydro_70; 1.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 34
FT CHAIN 35 1476
FT DOMAIN 35 1051
FT DOMAIN 1097 1476
FT REPEAT 1097 1130
FT DOMAIN 1161 1470
FT REPEAT 1161 1210
FT REPEAT 1225 1275
FT REPEAT 1290 1340
FT REPEAT 1355 1405
FT REPEAT 1420 1470
FT VARIANT 62 65
FT VARIANT 65 65
FT VARIANT 68 68
FT VARIANT 78 78
FT VARIANT 86 86
FT VARIANT 89 89
FT VARIANT 168 168
FT VARIANT 276 276
FT VARIANT 399 399
FT VARIANT 474 474
FT VARIANT 512 512
FT VARIANT 519 519
FT VARIANT 701 701
FT VARIANT 708 708
FT VARIANT 938 938
FT VARIANT 952 957
FT VARIANT 963 964
FT VARIANT 968 970
FT VARIANT 1086 1086
FT VARIANT 1158 1158
FT VARIANT 1163 1163
FT VARIANT 1168 1168
FT VARIANT 1182 1182
FT VARIANT 1234 1234
FT VARIANT 1263 1263
FT VARIANT 1263 1263
FT VARIANT 1264 1264
FT VARIANT 1272 1272
FT VARIANT 1329 1329
FT VARIANT 1394 1394
FT VARIANT 1402 1402
FT VARIANT 1459 1459
FT CONFLICT 570 570
FT CONFLICT 800 817
FT CONFLICT 1310 1310
FT CONFLICT 1476 AA; 165685 MW; 3479B62B07694D98 CRC64;
SQ SEQUENCE 1476 AA; 165685 MW; 3479B62B07694D98 CRC64;
Query Match 5.9%; Score 88; DB 1; Length 1476;
Best Local Similarity 19.3%; Pred. No. 71;
Matches 60; Conservative 51; Mismatches 122; Indels 78; Gaps 14;
Qy 35 KYGLYISGQYKPSVSVFSNFSVKE-----TNVITKNILAKKVDSTETKTDSVG 86
Db 992 KYPELFARKQISTGVPMDPSVKIKQWSAKYPNGNIGLRGAGYVLKQDQ---ATNTYFNIS 1048
Qy 87 ISNPSNFTIPYTAVFQDNSVNFN-----GTIGYTFABGTRVEIGSVYEEFDVKPNP 137
```

```
Db 1049 DNKEINF-LPKTLNQDSQVGFSDGKGYVYYSGYO-AKNTFISEGDKWYFD---NN 1103
QY 138 GYTLSDA-----YRYFALAREMKGNSFTPKKVSNSIEHTVMRNDGL-----SIISVIVNV 188
Db 1104 GYMTGAQSIINGVNYFSLNGLQLRDAILKNEGTYY---GNDGRYENGYYQFMSGV 1160
QY 189 CYDFSLNLSISPYICGGAGVDAIEFPDLHIKFAYOSKLGIAYSLPSNISLFLASLYYHK 248
Db 1161 WRHFNGEMSVGLTVIDG---QVQYFD---EMGYQAKGKFVTADGKIR-----YFDK 1207
QY 249 VMGNQFKN-----LNVOHVAELASIPKITSAVATLN---I 280
Db 1208 QSGNMYRNRFTIENEBGKWLILGEDCAAVTGSQTINGOHLYPFRANGVQVKGEFVTDHGR 1267
QY 281 GYFGGEIGARL 291
Db 1268 SYDGNSGDQI 1278
```

Search completed: July 8, 2003, 10:18:14  
Job time : 26 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 10:16:30 ; Search time 26 Seconds  
(without alignments)  
331.574 Million cell updates/sec

Title: US-10-062-624-40  
Perfect score: 1496  
Sequence: 1 MNKKLFTIINTVLVCLLSL.....AVATLNIGYFGGIGARLTF 293

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgm2\_6/ptodata/1/iaa/5A.COMB.pcp.\*  
2: /cgm2\_6/ptodata/1/iaa/5B.COMB.pcp.\*  
3: /cgm2\_6/ptodata/1/iaa/6A.COMB.pcp.\*  
4: /cgm2\_6/ptodata/1/iaa/6B.COMB.pcp.\*  
5: /cgm2\_6/ptodata/1/iaa/PTCUS.COMB.pcp.\*  
6: /cgm2\_6/ptodata/1/iaa/backfiles.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1496	100.0	293	4	US-09-660-587-40
2	498	33.3	283	4	US-09-660-587-10
3	498	33.3	283	4	US-09-261-358A-10
4	498	33.3	283	4	US-09-201-458-6
5	490.5	32.8	280	4	US-09-660-587-42
6	470.5	31.5	280	4	US-09-660-587-14
7	470.5	31.5	280	4	US-09-261-358A-14
8	470.5	31.5	280	4	US-09-201-458-10
9	468.5	31.3	280	4	US-08-953-326-17
10	459	30.7	287	3	US-08-733-230-2
11	459	30.7	287	4	US-08-953-326-2
12	456.5	30.5	284	4	US-09-660-587-15
13	456.5	30.5	284	4	US-09-261-358A-15
14	456.5	30.5	284	4	US-09-201-458-11
15	450.5	30.1	280	4	US-09-660-587-11
16	450.5	30.1	280	4	US-09-261-358A-11
17	450.5	30.1	280	4	US-09-201-458-7
18	446.5	29.8	280	4	US-09-660-587-6
19	446.5	29.8	280	4	US-09-261-358A-6
20	441.5	29.5	278	4	US-08-953-326-16
21	441.5	29.5	278	4	US-09-660-587-13
22	441.5	29.5	278	4	US-09-261-358A-13
23	441.5	29.5	278	4	US-09-201-458-9
24	441.5	29.5	283	4	US-09-660-587-4
25	441.5	29.5	283	4	US-09-261-358A-4
26	440.5	29.4	278	4	US-09-660-587-2
27	440.5	29.4	278	4	US-09-261-358A-2

28	440.5	29.4	278	4	US-09-201-458-2	Sequence 2, Appli
29	437	29.2	271	4	US-09-660-587-46	Sequence 46, Appli
30	434	29.0	281	4	US-09-660-587-9	Sequence 9, Appli
31	434	29.0	281	4	US-09-261-358A-9	Sequence 9, Appli
32	434	29.0	281	4	US-09-201-458-5	Sequence 5, Appli
33	427	28.5	286	4	US-08-953-326-15	Sequence 15, Appli
34	427	28.5	286	4	US-09-660-587-12	Sequence 12, Appli
35	427	28.5	286	4	US-09-261-358A-12	Sequence 12, Appli
36	427	28.5	286	4	US-09-201-458-8	Sequence 8, Appli
37	426	28.5	287	4	US-08-953-326-19	Sequence 19, Appli
38	426	28.5	287	4	US-09-660-587-8	Sequence 8, Appli
39	426	28.5	287	4	US-09-261-358A-8	Sequence 8, Appli
40	426	28.5	287	4	US-09-201-458-4	Sequence 4, Appli
41	423.5	28.3	276	4	US-09-660-587-44	Sequence 44, Appli
42	422	28.2	276	4	US-08-953-326-18	Sequence 18, Appli
43	409.5	27.4	280	3	US-08-733-230-4	Sequence 4, Appli
44	409.5	27.4	280	4	US-08-953-326-4	Sequence 4, Appli
45	294	19.7	282	3	US-08-733-230-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1  
US-09-660-587-40  
; Sequence 40, Application US/09660587  
; Patent No. 6392023  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Vere W.  
; APPLICANT: Yu, Xuejie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: US/09/660,587  
; CURRENT FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 40  
; LENGTH: 293  
; TYPE: PRT.  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of E. canis p28-1 protein  
US-09-660-587-40

Query Match	100.0%;	Score 1496;	DB 4;	Length 293;
Best Local Similarity	100.0%;	Pred. No. 3.3e-145;		
Matches 293;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MNKKLFTIINTVLVCLLSLPNISSSKAINNNKYYGLYISGQYKPSVSVFSPNPKET 60		
Db	1	MNKKLFTIINTVLVCLLSLPNISSSKAINNNKYYGLYISGQYKPSVSVFSPNPKET 60		
Qy	61	NVITKNLIALKQVDSIETKTDSVGSINSPSNFTPTAVFQDQNSVNFNGTIGYFAEGT 120		
Db	61	NVITKNLIALKQVDSIETKTDSVGSINSPSNFTPTAVFQDQNSVNFNGTIGYFAEGT 120		
Qy	121	RVEIEGSEYEDVKNPCGYTSLDAYRYFALAREMKNSTPPEKVSNSIFHTVNRNDGLS 180		
Db	121	RVEIEGSEYEDVKNPCGYTSLDAYRYFALAREMKNSTPPEKVSNSIFHTVNRNDGLS 180		
Qy	181	IISIVNVNVCYDFSNLNLISPIYICGGAGVDAIEFFDLHFKAYQSKGLTAYSLPSNISL 240		
Db	181	IISIVNVNVCYDFSNLNLISPIYICGGAGVDAIEFFDLHFKAYQSKGLTAYSLPSNISL 240		
Qy	241	FASLYHKWGNQFKNLNVQHVLAELASIPKITSNAVATLNIGYFGGIGARLTF 293		
Db	241	FASLYHKWGNQFKNLNVQHVLAELASIPKITSNAVATLNIGYFGGIGARLTF 293		

RESULT 2

```
US-09-660-587-10
; Sequence 10, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 10
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1B
US-09-660-587-10

Query Match      33.3%; Score 498; DB 4; Length 283;
Best Local Similarity 36.1%; Pred. No. 6e-43;
Matches 112; Conservative 60; Mismatches 94; Indels 44; Gaps 11;

QY 1 MNKLKFTIINTVLVCLLSL-----PNISSKAINNNAKKYGLYISQYKPSVSVP 52
Db 1 MNYKKIF--VSSALISLMSILPYQSFADPVTSDTNDTGINDSRE---GFYISVKYNPSISHF 55
QY 53 SNFSVKETNV-----ITKNLIALKKDVSIEKTDASVGISNPSNFTPIPTAV-FQDNSV 106
Db 56 RKFSAEAPINGNTSITKKVFLKKDGD-----IAQSANFNRTDPALEFQNLI 104
QY 107 N-FNGTIGYTPAETRVEIEGSEYEEFDVKNPGGYTLS--DAYRYFALAREMKGNSFTPK 163
Db 105 SGFSGSIGYAM-DGPRIELEAAQKFDKPNNDNTNSGDYKYFGLSRE---DAIADKK 160
QY 164 KVSNSIFHTVMRNDGLSIISVINVCYDFSLNNLSISPYICGGAGVDAIEFFDVLHIKFA 223
Db 161 -----YVVLKNEGITFMSLMVNTCYDITAEGVFPFIPYACAGVADLINVKDFNLKFS 213
QY 224 YQSKLGIAYSLPSNISLFSALYYHKVMGNQFNKLVQHVLAELASIPKITSAVATLINIGYF 283
Db 214 YQKIGISYPTPEVSFAFIGGYHGVGNFNKIPVITPVVLEGAPQTTLSALVTIDTGYF 273
QY 284 GGEIGARLTF 293
Db 274 GGEVGVRTFF 283

RESULT 3
US-09-261-358A-10
; Sequence 10, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 10
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1B
US-09-261-358A-10

Query Match      33.3%; Score 498; DB 4; Length 283;
Best Local Similarity 36.1%; Pred. No. 6e-43;
Matches 112; Conservative 60; Mismatches 94; Indels 44; Gaps 11;

QY 1 MNKLKFTIINTVLVCLLSL-----PNISSKAINNNAKKYGLYISQYKPSVSVP 52
Db 1 MNYKKIF--VSSALISLMSILPYQSFADPVTSDTNDTGINDSRE---GFYISVKYNPSISHF 55
QY 53 SNFSVKETNV-----ITKNLIALKKDVSIEKTDASVGISNPSNFTPIPTAV-FQDNSV 106
Db 56 RKFSAEAPINGNTSITKKVFLKKDGD-----IAQSANFNRTDPALEFQNLI 104
QY 107 N-FNGTIGYTPAETRVEIEGSEYEEFDVKNPGGYTLS--DAYRYFALAREMKGNSFTPK 163
Db 105 SGFSGSIGYAM-DGPRIELEAAQKFDKPNNDNTNSGDYKYFGLSRE---DAIADKK 160
QY 164 KVSNSIFHTVMRNDGLSIISVINVCYDFSLNNLSISPYICGGAGVDAIEFFDVLHIKFA 223
Db 161 -----YVVLKNEGITFMSLMVNTCYDITAEGVFPFIPYACAGVADLINVKDFNLKFS 213
QY 224 YQSKLGIAYSLPSNISLFSALYYHKVMGNQFNKLVQHVLAELASIPKITSAVATLINIGYF 283
Db 214 YQKIGISYPTPEVSFAFIGGYHGVGNFNKIPVITPVVLEGAPQTTLSALVTIDTGYF 273
QY 284 GGEIGARLTF 293
Db 274 GGEVGVRTFF 283
```

```
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1B
US-09-261-358A-10

Query Match      33.3%; Score 498; DB 4; Length 283;
Best Local Similarity 36.1%; Pred. No. 6e-43;
Matches 112; Conservative 60; Mismatches 94; Indels 44; Gaps 11;

QY 1 MNKLKFTIINTVLVCLLSL-----PNISSKAINNNAKKYGLYISQYKPSVSVP 52
Db 1 MNYKKIF--VSSALISLMSILPYQSFADPVTSDTNDTGINDSRE---GFYISVKYNPSISHF 55
QY 53 SNFSVKETNV-----ITKNLIALKKDVSIEKTDASVGISNPSNFTPIPTAV-FQDNSV 106
Db 56 RKFSAEAPINGNTSITKKVFLKKDGD-----IAQSANFNRTDPALEFQNLI 104
QY 107 N-FNGTIGYTPAETRVEIEGSEYEEFDVKNPGGYTLS--DAYRYFALAREMKGNSFTPK 163
Db 105 SGFSGSIGYAM-DGPRIELEAAQKFDKPNNDNTNSGDYKYFGLSRE---DAIADKK 160
QY 164 KVSNSIFHTVMRNDGLSIISVINVCYDFSLNNLSISPYICGGAGVDAIEFFDVLHIKFA 223
Db 161 -----YVVLKNEGITFMSLMVNTCYDITAEGVFPFIPYACAGVADLINVKDFNLKFS 213
QY 224 YQSKLGIAYSLPSNISLFSALYYHKVMGNQFNKLVQHVLAELASIPKITSAVATLINIGYF 283
Db 214 YQKIGISYPTPEVSFAFIGGYHGVGNFNKIPVITPVVLEGAPQTTLSALVTIDTGYF 273
QY 284 GGEIGARLTF 293
Db 274 GGEVGVRTFF 283

RESULT 4
US-09-201-458-6
; Sequence 6, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; TITLE OF INVENTION: canis and Uses Thereof
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 6
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1B
US-09-201-458-6

Query Match      33.3%; Score 498; DB 4; Length 283;
Best Local Similarity 36.1%; Pred. No. 6e-43;
Matches 112; Conservative 60; Mismatches 94; Indels 44; Gaps 11;

QY 1 MNKLKFTIINTVLVCLLSL-----PNISSKAINNNAKKYGLYISQYKPSVSVP 52
Db 1 MNYKKIF--VSSALISLMSILPYQSFADPVTSDTNDTGINDSRE---GFYISVKYNPSISHF 55
QY 53 SNFSVKETNV-----ITKNLIALKKDVSIEKTDASVGISNPSNFTPIPTAV-FQDNSV 106
Db 56 RKFSAEAPINGNTSITKKVFLKKDGD-----IAQSANFNRTDPALEFQNLI 104
QY 107 N-FNGTIGYTPAETRVEIEGSEYEEFDVKNPGGYTLS--DAYRYFALAREMKGNSFTPK 163
Db 105 SGFSGSIGYAM-DGPRIELEAAQKFDKPNNDNTNSGDYKYFGLSRE---DAIADKK 160
QY 164 KVSNSIFHTVMRNDGLSIISVINVCYDFSLNNLSISPYICGGAGVDAIEFFDVLHIKFA 223
Db 161 -----YVVLKNEGITFMSLMVNTCYDITAEGVFPFIPYACAGVADLINVKDFNLKFS 213
```





Db 1 MNCK-KP-FITTLVSLMSFLPGISFSDAVQND-NVGNFYISGKYVPSVSHFGVFSKQ 57

Qy 60 TNVITKNLIAKKVDISTETKTDASVGSINPSN-FTIP-YTAVFQDNS-VNFNGTIGYTF 116

Db 58 ERNTTIGVGLKQWDG-----STISKNSPENTFNPNYSFKYENNPFLGFAGAVGY-L 110

Qy 117 AEGTRVEIEGSEEDVDKNPGGYTLDAYRYFALAREMKGNSFTPKKVSNS----IPHT 172

Db 111 MNGPRIELMSYETFDVKNQGNKYNDKHYALTHNSGG-----KLSNAGDKFVF-- 161

Qy 173 VMNDGLSIISVINVCYDFSLNNLSISPYICGAGVDAIEFFDVLHIKFAYQSKLGIAI 232

Db 162 -LKNEGLDISLMLNACYDVISEGIPFPYCAGVGTDLISMFPAINPKISYQKLGLSY 220

Qy 233 SLPSNISLPASLYHKMGQNFKNLVQHVLAELASIPKITS-----AVATLNIYFGGE 286

Db 221 SISPEASVFGGHFKVIGNEFRDIP-----AMIPSTLTGNHFTIVLSVCHFGVE 273

Qy 287 IGARLTF 293

Db 274 LGGRNF 280

## RESULT 8

US-09-201-458-10

; Sequence 10, Application US/09201458A

; Patent No. 6458942

; GENERAL INFORMATION:

; APPLICANT: Walker, David H.

; APPLICANT: McBride, Jere W.

; APPLICANT: Yu, Xue-Jie

; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia

; FILE REFERENCE: D6152

; CURRENT APPLICATION NUMBER: US/09/201,458A

; CURRENT FILING DATE: 1998-11-30

; NUMBER OF SEQ ID NOS: 21

; SEQ ID NO 10

; LENGTH: 280

; TYPE: PRT

; ORGANISM: Ehrlichia chaffeensis

; FEATURE:

; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F

US-09-201-458-10

Query Match 31.5%; Score 470.5; DB 4; Length 280;

Best Local Similarity 37.5%; Pred. No. 3.9e-40;

Matches 115; Conservative 57; Mismatches 94; Indels 41; Gaps 14;

Qy 1 MNCK-KP-FITTLVSLMSFLPGISFSDAVQND-NVGNFYISGKYVPSVSHFGVFSKQ 59

Db 1 MNCK-KP-FITTLVSLMSFLPGISFSDAVQND-NVGNFYISGKYVPSVSHFGVFSKQ 57

Qy 60 TNVITKNLIAKKVDISTETKTDASVGSINPSN-FTIP-YTAVFQDNS-VNFNGTIGYTF 116

Db 58 ERNTTIGVGLKQWDG-----STISKNSPENTFNPNYSFKYENNPFLGFAGAVGY-L 110

Qy 117 AEGTRVEIEGSEEDVDKNPGGYTLDAYRYFALAREMKGNSFTPKKVSNS----IPHT 172

Db 111 MNGPRIELMSYETFDVKNQGNKYNDKHYALTHNSGG-----KLSNAGDKFVF-- 161

Qy 173 VMNDGLSIISVINVCYDFSLNNLSISPYICGAGVDAIEFFDVLHIKFAYQSKLGIAI 232

Db 162 -LKNEGLDISLMLNACYDVISEGIPFPYCAGVGTDLISMFPAINPKISYQKLGLSY 220

Qy 233 SLPSNISLPASLYHKMGQNFKNLVQHVLAELASIPKITS-----AVATLNIYFGGE 286

Db 221 SISPEASVFGGHFKVIGNEFRDIP-----AMIPSTLTGNHFTIVLSVCHFGVE 273

Qy 287 IGARLTF 293

Db 274 LGGRNF 280

## RESULT 9

US-08-953-326-17

; Sequence 17, Application US/08953326

; Patent No. 6251872

; GENERAL INFORMATION:

; APPLICANT: Barbet, Anthony F.

; APPLICANT: Ganta, Roman R.

; APPLICANT: McGuire, Travis C.

; APPLICANT: Burridge, Michael J.

; APPLICANT: Nyika, Aceme

; APPLICANT: Rurangirwa, Fred R.

; APPLICANT: Mahan, Suman M.

; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of

; FILE REFERENCE: UP-167C1

; CURRENT APPLICATION NUMBER: US/08/953,326

; CURRENT FILING DATE: 1997-10-17

; EARLIER APPLICATION NUMBER: 08/953,326

; EARLIER FILING DATE: 1997-10-17

; EARLIER APPLICATION NUMBER: 08/733,230

; EARLIER FILING DATE: 1996-10-17

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 17

; LENGTH: 280

; TYPE: PRT

; ORGANISM: Ehrlichia chaffeensis

US-08-953-326-17

Query Match 31.3%; Score 468.5; DB 4; Length 280;

Best Local Similarity 37.5%; Pred. No. 6.2e-40;

Matches 115; Conservative 57; Mismatches 94; Indels 41; Gaps 14;

Qy 1 MNCK-KP-FITTLVSLMSFLPGISFSDAVQND-NVGNFYISGKYVPSVSHFGVFSKQ 59

Db 1 MNCK-KP-FITTLVSLMSFLPGISFSDAVQND-NVGNFYISGKYVPSVSHFGVFSKQ 57

Qy 60 TNVITKNLIAKKVDISTETKTDASVGSINPSN-FTIP-YTAVFQDNS-VNFNGTIGYTF 116

Db 58 ERNTTIGVGLKQWDG-----STISKNSPENTFNPNYSFKYENNPFLGFAGAVGY-L 110

Qy 117 AEGTRVEIEGSEEDVDKNPGGYTLDAYRYFALAREMKGNSFTPKKVSNS----IPHT 172

Db 111 MNGPRIELMSYETFDVKNQGNKYNDKHYALTHNSGG-----KLSNAGDKFVF-- 161

Qy 173 VMNDGLSIISVINVCYDFSLNNLSISPYICGAGVDAIEFFDVLHIKFAYQSKLGIAI 232

Db 162 -LKNEGLDISLMLNACYDVISEGIPFPYCAGVGTDLISMFPAINPKISYQKLGLSY 220

Qy 233 SLPSNISLPASLYHKMGQNFKNLVQHVLAELASIPKITS-----AVATLNIYFGGE 286

Db 221 SISPEASVFGGHFKVIGNEFRDIP-----AMIPSTLTGNHFTIVLSVCHFGVE 273

Qy 287 IGARLTF 293

Db 274 LGGRNF 280

## RESULT 10

US-08-733-230-2

; Sequence 2, Application US/08733230

; Patent No. 6025338

; GENERAL INFORMATION:

; APPLICANT: Barbet, Anthony F.

; APPLICANT: Ganta, Roman Reddy

; APPLICANT: McGuire, Travis C.

; APPLICANT: Burridge, Michael J.

; APPLICANT: Nyika, Aceme

; APPLICANT: Rurangirwa, Fred R.

; APPLICANT: Mahan, Suman M.

; TITLE OF INVENTION: Nucleic Acid Vaccines Against

; TITLE OF INVENTION: Rickettsial Diseases and Methods of Use

RESULT 11  
US - 08-953-326-2  
; Sequence 2, Application US/08953326  
; Patent No. 6251872  
; GENERAL INFORMATION:  
; APPLICANT: Barbet, Anthony F.  
; APPLICANT: Ganta, Roman R.  
; APPLICANT: McGuire, Travis C.  
; APPLICANT: Burridge, Michael J.  
; APPLICANT: Nyika, Aceeme  
; APPLICANT: Rurangirwa, Fred R.  
; APPLICANT: Mahan, Suman M.

Query Match	30.5%;	Score 456.5;	DB 4;	Length 284;
Best Local Similarity	34.9%;	Pred. No. 1.1e-38;		
Matches 105; Conservative	58;	Mismatches 113;	Indels 25;	Gaps 8;

```
QY 1 MNKLKFTIINTVLVCLLS-LPNISSKAINNNAKYYGLYISGOYKPSVSFNSFVKE 59
DB 1 MNCKKIF--ITSTLSISLFLPGVSDVQIENNPGVSVISAKYMPMTASHFGKMSIKE 58

QY 60 TNVITKNLIALKKDVSITETKTDAISVGSINPNSFTIPTAVFQDNS-VNFGTIGYTFAE 118
DB 59 DSRDTKAVFGLKDWGKVTGPGTNSITFTEKDYFKY----ENNPFGLGAGAVGYSM-N 113

QY 119 GTRVEIEGSYEEDVKNPGGYTLDAYRYFALAREMKGNSFTPKBKVNSNIFHTVMRNDG 178
DB 114 GPRIEFVSYETFDVRNPGGNYKNDAMHYCALD-----TASSSTAGATTSSVMVKEN 165

QY 179 LSIISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDLVHLHKFAYQSKLGIAYSLPSNI 238
DB 166 LTDISLMLNACYDMLDGMFVSPYVCAGIGTDLVSVINATNPKLQYQKLGISYSINPEA 225

QY 239 SLFASLYHKVMGNQFNKLVQHV-----AELASIPKITSATVATLNIYFGGEIGARLT 292
DB 226 SIFIGGHRVIRVIGNEFKDIATSKVFTSSGNASSAVSPGFASAI--LDVCHFGIEIGGRFV 283

QY 293 F 293
DB 284 F 284

RESULT 13
US-09-261-358A-15
; Sequence 15, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; PRIOR FILING DATE: 1999-03-03
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 15
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
; FEATURE:
; OTHER INFORMATION: amino acid sequence of C. ruminantium MAP-1
US-09-261-358A-15

Query Match 30.5%; Score 456.5; DB 4; Length 284;
Best Local Similarity 34.9%; Pred. No. 1.1e-38;
Matches 105; Conservative 58; Mismatches 113; Indels 25; Gaps 8;

QY 1 MNKLKFTIINTVLVCLLS-LPNISSKAINNNAKYYGLYISGOYKPSVSFNSFVKE 59
DB 1 MNCKKIF--ITSTLSISLFLPGVSDVQIENNPGVSVISAKYMPMTASHFGKMSIKE 58

QY 60 TNVITKNLIALKKDVSITETKTDAISVGSINPNSFTIPTAVFQDNS-VNFGTIGYTFAE 118
DB 59 DSRDTKAVFGLKDWGKVTGPGTNSITFTEKDYFKY----ENNPFGLGAGAVGYSM-N 113

QY 119 GTRVEIEGSYEEDVKNPGGYTLDAYRYFALAREMKGNSFTPKBKVNSNIFHTVMRNDG 178
DB 114 GPRIEFVSYETFDVRNPGGNYKNDAMHYCALD-----TASSSTAGATTSSVMVKEN 165

QY 179 LSIISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDLVHLHKFAYQSKLGIAYSLPSNI 238
DB 166 LTDISLMLNACYDMLDGMFVSPYVCAGIGTDLVSVINATNPKLQYQKLGISYSINPEA 225

QY 239 SLFASLYHKVMGNQFNKLVQHV-----AELASIPKITSATVATLNIYFGGEIGARLT 292
DB 226 SIFIGGHRVIRVIGNEFKDIATSKVFTSSGNASSAVSPGFASAI--LDVCHFGIEIGGRFV 283

QY 293 F 293
DB 284 F 284

RESULT 14
US-09-201-458-11
; Sequence 11, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 11
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
; FEATURE:
; OTHER INFORMATION: amino acid sequence of C. ruminantium MAP-1
US-09-201-458-11

Query Match 30.5%; Score 456.5; DB 4; Length 284;
Best Local Similarity 34.9%; Pred. No. 1.1e-38;
Matches 105; Conservative 58; Mismatches 113; Indels 25; Gaps 8;

QY 1 MNKLKFTIINTVLVCLLS-LPNISSKAINNNAKYYGLYISGOYKPSVSFNSFVKE 59
DB 1 MNCKKIF--ITSTLSISLFLPGVSDVQIENNPGVSVISAKYMPMTASHFGKMSIKE 58

QY 60 TNVITKNLIALKKDVSITETKTDAISVGSINPNSFTIPTAVFQDNS-VNFGTIGYTFAE 118
DB 59 DSRDTKAVFGLKDWGKVTGPGTNSITFTEKDYFKY----ENNPFGLGAGAVGYSM-N 113

QY 119 GTRVEIEGSYEEDVKNPGGYTLDAYRYFALAREMKGNSFTPKBKVNSNIFHTVMRNDG 178
DB 114 GPRIEFVSYETFDVRNPGGNYKNDAMHYCALD-----TASSSTAGATTSSVMVKEN 165

QY 179 LSIISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDLVHLHKFAYQSKLGIAYSLPSNI 238
DB 166 LTDISLMLNACYDMLDGMFVSPYVCAGIGTDLVSVINATNPKLQYQKLGISYSINPEA 225

QY 239 SLFASLYHKVMGNQFNKLVQHV-----AELASIPKITSATVATLNIYFGGEIGARLT 292
DB 226 SIFIGGHRVIRVIGNEFKDIATSKVFTSSGNASSAVSPGFASAI--LDVCHFGIEIGGRFV 283

QY 293 F 293
DB 284 F 284

RESULT 15
US-09-660-587-11
; Sequence 11, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; PRIOR FILING DATE: 2000-09-12
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
```

```
; SEQ ID NO 11
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1C
US-09-660-587-11

Query Match      30.1%; Score 450.5; DB 4; Length 280;
Best Local Similarity 35.6%; Pred. No. 4.3e-38;
Matches 105; Conservative 56; Mismatches 117; Indels 17; Gaps 7;

Qy 1 MNKKLFTIINTVLVCLLSLPNISSKAINNNAKYYGLYISQYKPSVFSNFSVKET 60
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1 MNCK-KFFITTALALPMSFLPGLLSEPVQDDSVS-GNFIYISGYKMPSSASHFGVFSAKEE 58
61 NVITKNLIALKQDVDSITETKTASVGISNPSNFTIPYTAVFQDNS-VNFNGTIGYTFAG 119
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
59 KNPTVALYGLKQDWNGVSASSHADAFNNKG-----YsfKYENNPFLGFAGAGYSNG-G 112
Qy 120 TRVEIGSYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGL 179
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
113 PRIEFVSYETFDVKNQGGNYKNDARYCALDRKASSTNATASH-----YVLLKNEGL 165
Qy 180 SIISVIVNVCYDFSLNLSISPYICGGAGVDAIEFPDLVHLIKPAYQSKLGIAYSLPSNIS 239
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
166 LDISLMLNACYDVVSEGIPIFYICAGVGTDLISMFEAINPKISYQOGKGLSLSINPEAS 225
Qy 240 LFASLYYHKVMGNQFNKLN-VQHVLAELASIPKITSAVATNLNIGYFGGEIGARLTF 293
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
226 VFVGGHFKHVAGNEFRDITLKAFATPSSAATPDLATVTLVSCHFGVELGGRNPF 280

RESULT 17
US-09-201-458-7
; Sequence 7, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; TITLE OF INVENTION: canis and Uses Thereof
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 7
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1C
US-09-201-458-7

Query Match      30.1%; Score 450.5; DB 4; Length 280;
Best Local Similarity 35.6%; Pred. No. 4.3e-38;
Matches 105; Conservative 56; Mismatches 117; Indels 17; Gaps 7;

Qy 1 MNKKLFTIINTVLVCLLSLPNISSKAINNNAKYYGLYISQYKPSVFSNFSVKET 60
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1 MNCK-KFFITTALALPMSFLPGLLSEPVQDDSVS-GNFIYISGYKMPSSASHFGVFSAKEE 58
61 NVITKNLIALKQDVDSITETKTASVGISNPSNFTIPYTAVFQDNS-VNFNGTIGYTFAG 119
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
59 KNPTVALYGLKQDWNGVSASSHADAFNNKG-----YsfKYENNPFLGFAGAGYSNG-G 112
Qy 120 TRVEIGSYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGL 179
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
113 PRIEFVSYETFDVKNQGGNYKNDARYCALDRKASSTNATASH-----YVLLKNEGL 165
Qy 180 SIISVIVNVCYDFSLNLSISPYICGGAGVDAIEFPDLVHLIKPAYQSKLGIAYSLPSNIS 239
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
166 LDISLMLNACYDVVSEGIPIFYICAGVGTDLISMFEAINPKISYQOGKGLSLSINPEAS 225
Qy 240 LFASLYYHKVMGNQFNKLN-VQHVLAELASIPKITSAVATNLNIGYFGGEIGARLTF 293
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
226 VFVGGHFKHVAGNEFRDITLKAFATPSSAATPDLATVTLVSCHFGVELGGRNPF 280

RESULT 16
US-09-261-358A-11
; Sequence 11, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 11
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1C
US-09-261-358A-11

Query Match      30.1%; Score 450.5; DB 4; Length 280;
Best Local Similarity 35.6%; Pred. No. 4.3e-38;
Matches 105; Conservative 56; Mismatches 117; Indels 17; Gaps 7;

Qy 1 MNKKLFTIINTVLVCLLSLPNISSKAINNNAKYYGLYISQYKPSVFSNFSVKET 60
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1 MNCK-KFFITTALALPMSFLPGLLSEPVQDDSVS-GNFIYISGYKMPSSASHFGVFSAKEE 58
61 NVITKNLIALKQDVDSITETKTASVGISNPSNFTIPYTAVFQDNS-VNFNGTIGYTFAG 119
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
59 KNPTVALYGLKQDWNGVSASSHADAFNNKG-----YsfKYENNPFLGFAGAGYSNG-G 112
Qy 120 TRVEIGSYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGL 179
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
113 PRIEFVSYETFDVKNQGGNYKNDARYCALDRKASSTNATASH-----YVLLKNEGL 165
```

LENGTH: 280  
TYPE: PRT  
ORGANISM: Ehrlichia canis  
FEATURE:  
OTHER INFORMATION: amino acid sequence of p28-6 protein  
US-09-660-587-6

Query Match 29.8%; Score 446.5; DB 4; Length 280;  
Best Local Similarity 36.0%; Pred. No. 1.1e-37;  
Matches 105; Conservative 58; Mismatches 110; Indels 19; Gaps 10;  
QY 6 KFTIINTVLVCLL-SLPNISSSKANNNAKYGLYISQYKPSVSVFSNFSVKETNVT 64  
DB 4 KKLITLMSLMYAPSIPTDITQDDNTGSGF--YISGKYVPSVSHFGVFSKERNST 61  
QY 65 KNLIALKQDVSIETKTDAVGSINPSN-FTIP-YTAVFQDNS-VNPNGTIGYTFAGETR 121  
DB 62 VGVFLGKHQWNG-----GTISNSSPENIFTQNTSFYKYNPNPFLGFAIGYSMG-GPR 114  
QY 122 VEIEGSYEEDVKNPGGTYLSDAYRYFALAREMKNSTPPEKVSNSIFHTVMRNDGLSI 181  
DB 115 LEVLVETFDVKNQNNYKNGAHCALSHH---SSATSMSSASNKF--VFLKNEGLTD 169  
QY 182 ISVIVNVCYDLSNLSISPYICGAGVDAIEFFDVLHKKFAYQSKLGIAYSLPSNISLF 241  
DB 170 LSFMINACYDIIIEGMPSPYICAGVGTDVSMFEAINPKISYQKGLGLGISSEASVF 229  
QY 242 ASLYHKVMGNQKMLNVQHVLAELASIPKITSATVATLNIGYFGGIGARLTF 293  
DB 230 ICGHFRHVGNEFRDIPAM-VPSGSLNPENQFAIVTLNVCHFGIELGGRFNF 280

RESULT 19  
US-09-261-358A-6  
Sequence 6, Application US/09261358A  
Patent No. 6403780  
GENERAL INFORMATION:  
APPLICANT: Walker, David H.  
APPLICANT: McBride, Jere W.  
APPLICANT: Yu, Xue-Jie  
TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof  
FILE REFERENCE: D6152CIP  
CURRENT APPLICATION NUMBER: US/09/261,358A  
CURRENT FILING DATE: 1998-03-03  
PRIOR APPLICATION NUMBER: 09/201,458  
PRIOR FILING DATE: 1998-11-30  
NUMBER OF SEQ ID NOS: 33  
SEQ ID NO 6  
LENGTH: 280  
TYPE: PRT  
ORGANISM: Ehrlichia canis  
FEATURE:  
OTHER INFORMATION: amino acid sequence of Eca288A3 protein  
US-09-261-358A-6

Query Match 29.8%; Score 446.5; DB 4; Length 280;  
Best Local Similarity 36.0%; Pred. No. 1.1e-37;  
Matches 105; Conservative 58; Mismatches 110; Indels 19; Gaps 10;  
QY 6 KFTIINTVLVCLL-SLPNISSSKANNNAKYGLYISQYKPSVSVFSNFSVKETNVT 64  
DB 4 KKLITLMSLMYAPSIPTDITQDDNTGSGF--YISGKYVPSVSHFGVFSKERNST 61  
QY 65 KNLIALKQDVSIETKTDAVGSINPSN-FTIP-YTAVFQDNS-VNPNGTIGYTFAGETR 121  
DB 62 VGVFLGKHQWNG-----GTISNSSPENIFTQNTSFYKYNPNPFLGFAIGYSMG-GPR 114  
QY 122 VEIEGSYEEDVKNPGGTYLSDAYRYFALAREMKNSTPPEKVSNSIFHTVMRNDGLSI 181  
DB 115 LEVLVETFDVKNQNNYKNGAHCALSHH---SSATSMSSASNKF--VFLKNEGLTD 169  
QY 182 ISVIVNVCYDLSNLSISPYICGAGVDAIEFFDVLHKKFAYQSKLGIAYSLPSNISLF 241

DB 170 LSFMINACYDIIIEGMPSPYICAGVGTDVSMFEAINPKISYQKGLGLGISSEASVF 229  
QY 242 ASLYHKVMGNQKMLNVQHVLAELASIPKITSATVATLNIGYFGGIGARLTF 293  
DB 230 ICGHFRHVGNEFRDIPAM-VPSGSLNPENQFAIVTLNVCHFGIELGGRFNF 280

RESULT 20  
US-08-953-326-16  
Sequence 16, Application US/08953326  
Patent No. 6251872  
GENERAL INFORMATION:  
APPLICANT: Barbet, Anthony F.  
APPLICANT: Ganta, Roman R.  
APPLICANT: McGuire, Travis C.  
APPLICANT: Burridge, Michael J.  
APPLICANT: Nyika, Aceme  
APPLICANT: Rurangirwa, Fred R.  
APPLICANT: Mahan, Suman M.  
TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of  
TITLE OF INVENTION: Animals and Humans  
FILE REFERENCE: UF-167C1  
CURRENT APPLICATION NUMBER: US/08/953,326  
CURRENT FILING DATE: 1997-10-17  
EARLIER APPLICATION NUMBER: 08/953,326  
EARLIER FILING DATE: 1997-10-17  
EARLIER APPLICATION NUMBER: 08/733,230  
EARLIER FILING DATE: 1996-10-17  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 16  
LENGTH: 278  
TYPE: PRT  
ORGANISM: Ehrlichia chaffeensis  
US-08-953-326-16

Query Match 29.5%; Score 441.5; DB 4; Length 278;  
Best Local Similarity 37.4%; Pred. No. 3.6e-37;  
Matches 111; Conservative 56; Mismatches 107; Indels 23; Gaps 12;  
QY 1 MNKKFTIINTVLVCLLS-LPNISSSKAI--NNNAKYGLYISQYKPSVSVFSNFSVK 58  
DB 1 MNCK-KP-FITLALVSLMFLPGISFSDPVQGDNISGNF--YVSGKYWPSASHGMSAK 56  
QY 59 ETNVITKNLIALKQDVSIETKTDAVGSINPSNFTIPYTAVFQDNS-VNPNGTIGYTF 117  
DB 57 ECKNPTVALYGLKQWEGISSSSHNDNFNNKG-----YSPKYENNPFLGFAIGYSMG 111  
QY 118 EGTRVEIEGSYEEDVKNPGGTYLSDAYRYFALAREMKNSTPPEKVSNSIFHTVMRND 177  
DB 112 -GPRVEFEVSFETFDVKNQNNYKNDHRYCALGQQ--DNSGIPKTSK-----YVLLKSE 163  
QY 178 GLSIISVIVNVCYDLSNLSISPYICGAGVDAIEFFDVLHKKFAYQSKLGIAYSLPSN 237  
DB 164 GLDIISPLNACYDIINESPLSPYICAGVGTDLISMEATNPKISYQKGLGLSYINPE 223  
QY 238 ISLFASYIYHKVMGNQKMLNVQHVLAELASIPKITSATVATLNIGYFGGIGARLTF 293  
DB 224 ASVFIGHFKHVGNEFRDIPAM-VPSGSLNPENQFAIVTLNVCHFGIELGGRFNF 278

RESULT 21  
US-09-660-587-13  
Sequence 13, Application US/09660587  
Patent No. 6392023  
GENERAL INFORMATION:  
APPLICANT: Walker, David H.  
APPLICANT: McBride, Jere W.  
APPLICANT: Yu, Xue-Jie  
TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof  
FILE REFERENCE: D6152CIP2

```
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 13
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1E
US-09-660-587-13

Query Match      29.5%; Score 441.5; DB 4; Length 278;
Best Local Similarity 37.4%; Pred. No. 3.6e-37;
Matches 111; Conservative 56; Mismatches 107; Indels 23; Gaps 12;

Qy      1 MNKKLFTIINTVLVCLLS-LPNTSSSKAI--NNNAKKYGYLYISGOYKPSVVSFNSVK 58
Db      1 MNCK-KF-FITTALVSLMSFLPGISFSDPVQGDNISGNF--YVSGKYMPSASHFGMFSAK 56

Qy      59 ETNVITKNLIALKDVDSIEKTKDASVGSINPSNFTIPYTAVFQDNS--VNFNGTIGYTF 117
Db      57 EEKNPTVALYGLKQDWEGISSSHNDHFNKNG-----YSPKYENNPFPGAGAGYSMG 111

Qy      118 EGTREIEGSYEEDVKNPGGYTLSDAYRYPALAREMKGNSFTPKKVSNSIFHTVMRND 177
Db      112 -GPRVEFEVSJETPDVKNQGNKYNDHRYCALGQ--DNSGIPKTSK-----YVLLKSE 163

Qy      178 GLSIISVIVNVVDFSLNNLSISPYICGGAGVDAIEFPDVLHIKFAVQSKLGIAYSLSN 237
Db      164 GLLDISFPLNACYDIINESIPLSPYICAGVGTDLISMFEATNPKISYOGKLGLSYNPE 223

Qy      238 ISLFASYHYHKVMGNQFNKL--NVQHVLAELASIPKITSATVATLNIGYFGGEIGARLTF 293
Db      224 ASVFIGGHFKVIGNEFRDPTLKAFVTSSATPDL--AIVTSLVCHFGIELGGRFNP 278

RESULT 22
US-09-261-358A-13
; Sequence 13, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 13
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1E
US-09-261-358A-13

Query Match      29.5%; Score 441.5; DB 4; Length 278;
Best Local Similarity 37.4%; Pred. No. 3.6e-37;
Matches 111; Conservative 56; Mismatches 107; Indels 23; Gaps 12;

Qy      1 MNKKLFTIINTVLVCLLS-LPNTSSSKAI--NNNAKKYGYLYISGOYKPSVVSFNSVK 58
Db      1 MNCK-KF-FITTALVSLMSFLPGISFSDPVQGDNISGNF--YVSGKYMPSASHFGMFSAK 56

Qy      59 ETNVITKNLIALKDVDSIEKTKDASVGSINPSNFTIPYTAVFQDNS--VNFNGTIGYTF 117
Db      57 EEKNPTVALYGLKQDWEGISSSHNDHFNKNG-----YSPKYENNPFPGAGAGYSMG 111

Qy      118 EGTREIEGSYEEDVKNPGGYTLSDAYRYPALAREMKGNSFTPKKVSNSIFHTVMRND 177
Db      112 -GPRVEFEVSJETPDVKNQGNKYNDHRYCALGQ--DNSGIPKTSK-----YVLLKSE 163

Qy      178 GLSIISVIVNVVDFSLNNLSISPYICGGAGVDAIEFPDVLHIKFAVQSKLGIAYSLSN 237
Db      164 GLLDISFPLNACYDIINESIPLSPYICAGVGTDLISMFEATNPKISYOGKLGLSYNPE 223

Qy      238 ISLFASYHYHKVMGNQFNKL--NVQHVLAELASIPKITSATVATLNIGYFGGEIGARLTF 293
Db      224 ASVFIGGHFKVIGNEFRDPTLKAFVTSSATPDL--AIVTSLVCHFGIELGGRFNP 278

RESULT 23
US-09-201-458-9
; Sequence 9, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 9
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1E
US-09-201-458-9

Query Match      29.5%; Score 441.5; DB 4; Length 278;
Best Local Similarity 37.4%; Pred. No. 3.6e-37;
Matches 111; Conservative 56; Mismatches 107; Indels 23; Gaps 12;

Qy      1 MNKKLFTIINTVLVCLLS-LPNTSSSKAI--NNNAKKYGYLYISGOYKPSVVSFNSVK 58
Db      1 MNCK-KF-FITTALVSLMSFLPGISFSDPVQGDNISGNF--YVSGKYMPSASHFGMFSAK 56

Qy      59 ETNVITKNLIALKDVDSIEKTKDASVGSINPSNFTIPYTAVFQDNS--VNFNGTIGYTF 117
Db      57 EEKNPTVALYGLKQDWEGISSSHNDHFNKNG-----YSPKYENNPFPGAGAGYSMG 111

Qy      118 EGTREIEGSYEEDVKNPGGYTLSDAYRYPALAREMKGNSFTPKKVSNSIFHTVMRND 177
Db      112 -GPRVEFEVSJETPDVKNQGNKYNDHRYCALGQ--DNSGIPKTSK-----YVLLKSE 163

Qy      178 GLSIISVIVNVVDFSLNNLSISPYICGGAGVDAIEFPDVLHIKFAVQSKLGIAYSLSN 237
Db      164 GLLDISFPLNACYDIINESIPLSPYICAGVGTDLISMFEATNPKISYOGKLGLSYNPE 223

Qy      238 ISLFASYHYHKVMGNQFNKL--NVQHVLAELASIPKITSATVATLNIGYFGGEIGARLTF 293
Db      224 ASVFIGGHFKVIGNEFRDPTLKAFVTSSATPDL--AIVTSLVCHFGIELGGRFNP 278

RESULT 24
US-09-660-587-4
; Sequence 4, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
```

; CURRENT FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 4  
; LENGTH: 283  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of p28-5 protein  
US-09-660-587-4

Query Match 29.5%; Score 441; DB 4; Length 283;  
Best Local Similarity 36.9%; Pred. No. 4.1e-37;  
Matches 109; Conservative 46; Mismatches 126; Indels 14; Gaps 7;  
QY 1 MNKLKFTIINTVLVCLSLPNISSKAINNAKKYGLYISGOYKPSVFSNFSVKET 60  
DB 1 MNCKKVFT-ISALISSIYFLPNVSYNPNVYGN-SYGNFYISGKYMPSVPHGIFSAEE 58  
QY 61 NVITKNLALKDVSIEFTKTDASVIGSNPSNFTI-PYTAVFQDNS-VNFGTIGYTTAE 118  
DB 59 KKKTTVVYGLKEN-----WAGDAISSQSPDDNFTIRNYSFKYASKNKFLGFAVAIGYSIG- 112  
QY 119 GTRVEIGSYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKEKVSNIFFHTVMRNDG 178  
DB 113 SPRIEVSMEYAFDVKNPNGDNYKAYRYCALSHQDDADD----DMTSATDKFYVLINEG 168  
QY 179 LSIISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDVLHIKPAYQSKLGIAYSLPSNI 238  
DB 169 LLNISFMTNICYETASKNIPLSPYICAGIGTDLIHFETHPKISYQKGLAYFVSAES 228  
QY 239 SLFASLYHKVMGNQFKNLVQHVLAELASIPKITSATVATNLNIGYFGGIGARLTF 293  
DB 229 SVSFGIYFHKIINNKFNKVPAMVPINSDIIVGPQFATVTLNVCYFGLGCRFNF 283

RESULT 25  
US-09-261-358A-4  
; Sequence 4, Application US/09261358A  
; Patent No. 6403780  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof  
; FILE REFERENCE: D6152CIP  
; CURRENT APPLICATION NUMBER: US/09/261,358A  
; CURRENT FILING DATE: 1999-03-03  
; PRIOR APPLICATION NUMBER: 09/201,458  
; PRIOR FILING DATE: 1998-11-30  
; NUMBER OF SEQ ID NOS: 33  
; SEQ ID NO 4  
; LENGTH: 283  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of EC28SA2 protein  
US-09-261-358A-4

Query Match 29.5%; Score 441; DB 4; Length 283;  
Best Local Similarity 36.9%; Pred. No. 4.1e-37;  
Matches 109; Conservative 46; Mismatches 126; Indels 14; Gaps 7;  
QY 1 MNKLKFTIINTVLVCLSLPNISSKAINNAKKYGLYISGOYKPSVFSNFSVKET 60  
DB 1 MNCKKVFT-ISALISSIYFLPNVSYNPNVYGN-SYGNFYISGKYMPSVPHGIFSAEE 58  
QY 61 NVITKNLALKDVSIEFTKTDASVIGSNPSNFTI-PYTAVFQDNS-VNFGTIGYTTAE 118  
DB 59 KKKTTVVYGLKEN-----WAGDAISSQSPDDNFTIRNYSFKYASKNKFLGFAVAIGYSIG- 112

QY 119 GTRVEIGSYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKEKVSNIFFHTVMRNDG 178  
DB 113 SPRIEVSMEYAFDVKNPNGDNYKAYRYCALSHQDDADD----DMTSATDKFYVLINEG 168  
QY 179 LSIISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDVLHIKPAYQSKLGIAYSLPSNI 238  
DB 169 LLNISFMTNICYETASKNIPLSPYICAGIGTDLIHFETHPKISYQKGLAYFVSAES 228  
QY 239 SLFASLYHKVMGNQFKNLVQHVLAELASIPKITSATVATNLNIGYFGGIGARLTF 293  
DB 229 SVSFGIYFHKIINNKFNKVPAMVPINSDIIVGPQFATVTLNVCYFGLGCRFNF 283

RESULT 26  
US-09-660-587-2  
; Sequence 2, Application US/09660587  
; Patent No. 6392023  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof  
; FILE REFERENCE: D6152CIP  
; CURRENT APPLICATION NUMBER: US/09/660,587  
; CURRENT FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 2  
; LENGTH: 278  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of E. canis p28-7 protein  
US-09-660-587-2

Query Match 29.4%; Score 440.5; DB 4; Length 278;  
Best Local Similarity 35.6%; Pred. No. 4.5e-37;  
Matches 105; Conservative 56; Mismatches 107; Indels 27; Gaps 8;  
QY 6 KFTIINTVLVCLL-SLPNISSKAINNAKKYGLYISGOYKPSVFSNFSVKETNVT 64  
DB 4 KKILITLISLWYISPSISFSDTI-QDGNMGNFYISGKYPVSVSHGFSFAKEESKST 62  
QY 65 KNLIALKDVDSIETKTDDASVIGSNPSNFTIP-YTAVFQDNS-VNFGTIGYTTAEGRV 122  
DB 63 VGVFGLKHDWDG-----SPILKNKHADFTVPNYSFRYNNPFLGFAIGYSMG-GPRI 115  
QY 123 BIEGSEYEEFDVKNPGGYTLSDAYRYFALARE---MKGNSFTPKKEKVSNIFFHTVMRNDG 178  
DB 116 EFEISYEAFDVKSPNINYNQNDHRYCALSHHTSAAEADKF-----VFLKNEG 163  
QY 179 LSIISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDVLHIKPAYQSKLGIAYSLPSNI 238  
DB 164 LIDISLAINACYDIINDKVPVSYICAGIGTDLISMFEATSPKISYQKGLISYINPET 223  
QY 239 SLFASLYHKVMGNQFKNLVQHVLAELASIPKITSATVATNLNIGYFGGIGARLTF 293  
DB 224 SVFICGGHFRIGNEFRDIPAIVPSNNTTISGQFATVTLNVCYFGLGCRFNF 278

RESULT 27  
US-09-261-358A-2  
; Sequence 2, Application US/09261358A  
; Patent No. 6403780  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof  
; FILE REFERENCE: D6152CIP  
US-09-261-358A-2



```
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 2
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of Eca28-1 protein
US-09-261-358A-2

Query Match      29.4%; Score 440.5; DB 4; Length 278;
Best Local Similarity 35.6%; Pred. No. 4.5e-37;
Matches 105; Conservative 56; Mismatches 107; Indels 27; Gaps 8;

Qy      6 KFTIINTVLVCLL-SLPNISSKAINNNAKKYGLYISGQKPSVSVFNSFKETNVT 64
Db      4 KKIILITLISLWYISFISFSDTI-QDGNMGNFYISGKVPVSUHFSGFSAKESKST 62

Qy      65 KNLIALKKDVSIEKTKDASVIGSNPSNFTIP-YTAVFQDMS-VNFMGTIGYTFAGTRV 122
Db      63 VGVFLKHDWDG-----SPILKNKHADFTVPNYSFRYENNPFLGFAGAGYSMG-GPRI 115

Qy      123 EIEGSYEEDVKVPGGYTSLDAYRYPALARE----MKGNSFTPKKVSNSIFHTVMRNDG 178
Db      116 EFEISYEAFDVKSPNINYQNDAHRYCALSHHTSAAMEADKF-----VFLKNEG 163

Qy      179 LSIISVIVNVVYDFSLNLSISPYICGGAGVDAREFFDLVHLIKFAYOSKLGIAVSLPSNI 238
Db      164 LIDISLAINACYDIINDKVPVSPICAGIGTDLISMFEATSPKISYQCKLGISINPET 223

Qy      239 SLFASLYHKVMGNQFNKLVNQHVAELASIPKITSAVATNLNIGYFGGEIGARLTF 293
Db      224 SVFIGGHFRIIGNEFRDIPALVPSNSTTISGPOFATVTLNVCHFLGELGGRNF 278

RESULT 28
US-09-201-458-2
; Sequence 2, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 2
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of a 30 kDa immunoreactive
; OTHER INFORMATION: protein of Ehrlichia canis
US-09-201-458-2

Query Match      29.4%; Score 440.5; DB 4; Length 278;
Best Local Similarity 35.6%; Pred. No. 4.5e-37;
Matches 105; Conservative 56; Mismatches 107; Indels 27; Gaps 8;

Qy      6 KFTIINTVLVCLL-SLPNISSKAINNNAKKYGLYISGQKPSVSVFNSFKETNVT 64
Db      4 KKIILITLISLWYISFISFSDTI-QDGNMGNFYISGKVPVSUHFSGFSAKESKST 62

Qy      65 KNLIALKKDVSIEKTKDASVIGSNPSNFTIP-YTAVFQDMS-VNFMGTIGYTFAGTRV 122
Db      63 VGVFLKHDWDG-----SPILKNKHADFTVPNYSFRYENNPFLGFAGAGYSMG-GPRI 115

Qy      123 EIEGSYEEDVKVPGGYTSLDAYRYPALARE----MKGNSFTPKKVSNSIFHTVMRNDG 178
Db      116 EFEISYEAFDVKSPNINYQNDAHRYCALSHHTSAAMEADKF-----VFLKNEG 163

Qy      179 LSIISVIVNVVYDFSLNLSISPYICGGAGVDAREFFDLVHLIKFAYOSKLGIAVSLPSNI 238
Db      164 LIDISLAINACYDIINDKVPVSPICAGIGTDLISMFEATSPKISYQCKLGISINPET 223

Qy      239 SLFASLYHKVMGNQFNKLVNQHVAELASIPKITSAVATNLNIGYFGGEIGARLTF 293
Db      224 SVFIGGHFRIIGNEFRDIPALVPSNSTTISGPOFATVTLNVCHFLGELGGRNF 278

US-09-660-587-46
; Sequence 9, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 45
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis p28-9 protein
US-09-660-587-46

Query Match      29.2%; Score 437; DB 4; Length 271;
Best Local Similarity 37.4%; Pred. No. 1e-36;
Matches 110; Conservative 44; Mismatches 110; Indels 30; Gaps 8;

Qy      1 MNKLKFTIINTVLVCLLSLPNISSKAINNNAKKYGLYISGQKPSVSVFNSFKET 60
Db      1 MNYK-RFVVGVTLSFTVFVFLSDGAFSDA--NFSEGRRLYIGSYQKVGIPNFSNFAET 57

Qy      61 -NVITKNLIALKKDVSIEKTKDASVIGSNPSNFTIPYTAVFQDMSVNFMGTIGYTFAG 119
Db      58 IPGITKKIPALGLDKSEINTH-----SNFTSYDPTYASSFAGFGIGIY-YVND 106

Qy      120 TRVEIEGSYEEDVKVPGGYTSLDAYRYPALARE--KGNSTPKKVSNSIFHTVMRND 177
Db      107 FRVEFGSYENFEPERQWYPENSQSYKFFALSRNATNSDNKF-----IVLENN 154

Qy      178 GLSIISVIVNVVYDFSLNLSISPYICGGAGVDAREFFDLVHLIKFAYOSKLGIAVSLPSN 237
Db      155 GVVDKSLNVNVCYDIASGSIPLAPYMCAGVADYIKFLGSLPKFSQVRFVGNYPNVN 214

Qy      238 ISLPASLYHKVMGNQFNKLVN-QHVAELASIPKITSAVATNLNIGYFGGEIGAR 290
Db      215 TMLFGGYYHKVVDGRHERVEIAHYHTALSDVPTTTSASATLNTDIFGWEIGFR 268

RESULT 30
US-09-660-587-9
; Sequence 9, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
```

```
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 9
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis p28
US-09-660-587-9

Query Match      29.0%; Score 434; DB 4; Length 281;
Best Local Similarity 35.6%; Pred. No. 2.1e-36;
Matches 109; Conservative 57; Mismatches 102; Indels 38; Gaps 13;

QY 1 MNKLKFTIINTVLVCLL-SLPNISS-----KAINNNAKKYGLYISGOYKPSVSFNSF 55
DB 1 MNYKKVF--ITSALISLSSLPGVSFSDPAGSGINGN-----FYISGYMPSASHFGVF 52

QY 56 SVKETNVTIKNLIALKXDVDSIETKTDSVGISNPSN-FTIP-YTAVFQDNS-VNFNGTI 112
DB 53 SAKERNTTVGVFGLKQWDG-----SAISNSSPNDVFTVSNYSFKYENNPFLGFAGAI 106

QY 113 GYTFAGETRVEIEGSEYEFVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHT 172
DB 107 GYSM-DGPRIELEVSJETFDVKNQNNYKNEAHRYCALSH-----NSAADMSASNNF--V 159

QY 173 VMRNDGLSIISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDLVHLKFAVQSKLGAIY 232
DB 160 FLKNEGLLDISFMLNACVDVVGEGIPFSPYICAGIGTDLVSMFEATNPKISYQKGLGSLY 219

QY 233 SLPSNISLFASYHHKVMGNQFNKLV-----QHVLAELASIPKITSATVATLNIGYFGGEI 287
DB 220 SISPEASVFIGGHPKHVIGNEFRDIPTIPTGSTLAGKNYP-----AIVILDVCHFGIEL 275

QY 288 GARLTF 293
DB 276 GGRFAF 281

RESULT 31
US-09-261-358A-9
; Sequence 9, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 9
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis p28
US-09-261-358A-9

Query Match      29.0%; Score 434; DB 4; Length 281;
Best Local Similarity 35.6%; Pred. No. 2.1e-36;
Matches 109; Conservative 57; Mismatches 102; Indels 38; Gaps 13;

QY 1 MNKLKFTIINTVLVCLL-SLPNISS-----KAINNNAKKYGLYISGOYKPSVSFNSF 55
DB 1 MNYKKVF--ITSALISLSSLPGVSFSDPAGSGINGN-----FYISGYMPSASHFGVF 52

QY 56 SVKETNVTIKNLIALKXDVDSIETKTDSVGISNPSN-FTIP-YTAVFQDNS-VNFNGTI 112
DB 53 SAKERNTTVGVFGLKQWDG-----SAISNSSPNDVFTVSNYSFKYENNPFLGFAGAI 106

QY 113 GYTFAGETRVEIEGSEYEFVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHT 172
DB 107 GYSM-DGPRIELEVSJETFDVKNQNNYKNEAHRYCALSH-----NSAADMSASNNF--V 159

QY 173 VMRNDGLSIISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDLVHLKFAVQSKLGAIY 232
DB 160 FLKNEGLLDISFMLNACVDVVGEGIPFSPYICAGIGTDLVSMFEATNPKISYQKGLGSLY 219

QY 233 SLPSNISLFASYHHKVMGNQFNKLV-----QHVLAELASIPKITSATVATLNIGYFGGEI 287
DB 220 SISPEASVFIGGHPKHVIGNEFRDIPTIPTGSTLAGKNYP-----AIVILDVCHFGIEL 275

QY 288 GARLTF 293
DB 276 GGRFAF 281

RESULT 32
US-09-201-458-5
; Sequence 5, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 5
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis p28
US-09-201-458-5

Query Match      29.0%; Score 434; DB 4; Length 281;
Best Local Similarity 35.6%; Pred. No. 2.1e-36;
Matches 109; Conservative 57; Mismatches 102; Indels 38; Gaps 13;

QY 1 MNKLKFTIINTVLVCLL-SLPNISS-----KAINNNAKKYGLYISGOYKPSVSFNSF 55
DB 1 MNYKKVF--ITSALISLSSLPGVSFSDPAGSGINGN-----FYISGYMPSASHFGVF 52

QY 56 SVKETNVTIKNLIALKXDVDSIETKTDSVGISNPSN-FTIP-YTAVFQDNS-VNFNGTI 112
DB 53 SAKERNTTVGVFGLKQWDG-----SAISNSSPNDVFTVSNYSFKYENNPFLGFAGAI 106

QY 113 GYTFAGETRVEIEGSEYEFVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHT 172
DB 107 GYSM-DGPRIELEVSJETFDVKNQNNYKNEAHRYCALSH-----NSAADMSASNNF--V 159

QY 173 VMRNDGLSIISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDLVHLKFAVQSKLGAIY 232
DB 160 FLKNEGLLDISFMLNACVDVVGEGIPFSPYICAGIGTDLVSMFEATNPKISYQKGLGSLY 219

QY 233 SLPSNISLFASYHHKVMGNQFNKLV-----QHVLAELASIPKITSATVATLNIGYFGGEI 287
DB 220 SISPEASVFIGGHPKHVIGNEFRDIPTIPTGSTLAGKNYP-----AIVILDVCHFGIEL 275

QY 288 GARLTF 293
DB 276 GGRFAF 281
```





Qy	113	GUTFAEGTRVEIEGSEYEFEDVKNPGGYTLSDAIVYFALARE----	MKGSN-----FTPKVK	165
Db	107	GYSIG-NSRIELEVSHEIFDTKNPGNNLYNDHSKYCALSHGSHICSGDN	SGDWYTAKTDK	165
Qy	166	SNSIFHTVMRNDGLSIIISVINVCYDFSLNLSPICYCGAGVDIAEP	FVDLHIKFAVQ	225
Db	166	-----FVLLKNEGLLDVSFMLNACVDITTEKMPPSPYICAGITGLI	SMFPETTONKSIVQ	220
Qy	226	SKLGTAISLPNISILFASLYTHKWGNQFK-----	NLVQHVAELASIPKITS	275
Db	221	GKLGLNYTINRSVSVFAGGHFHKVIGNEFGIPTLLPDG	SNIKVQQ-----SAT	269
Qy	276	ATLNGYFGGEIGARLTFF	293	
Db	270	VTLDVCHFGLEIGSRFFF	287	

RESULT 39

```

US-09-261-358A-8
; Sequence 8, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 8
; LENGTH: 267
; TYPE: PRP
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of ECa28SA1 protien
US-09-261-358A-8

```

Search completed: July 8, 2003, 10:21:02  
Job time : 28 secs

```

: Sequence 4, Application US/09201458A
: Patent No. 6458942
: GENERAL INFORMATION:
: APPLICANT: Walker, David H.
: APPLICANT: McBride, Jere W.
: APPLICANT: Yu, Xue-die
: TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
: TITLE OF INVENTION: canis and Uses Thereof
: FILE REFERENCE: D6152
: CURRENT APPLICATION NUMBER: US/09/201,458A
: CURRENT FILING DATE: 1998-11-30
: NUMBER OF SEQ ID NOS: 21
: SEQ ID NO 4
: LENGTH: 287
: TYPE: PRT
: ORGANISM: Ehrlichia canis
: FEATURE:
: OTHER INFORMATION: amino acid sequence of E. canis 28-kDa protein-2
: OTHER INFORMATION: (partial)
: US-09-201-458-4

```

**RESULT 40**

US-09-201-458-4

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 4, 2003, 04:06:09 ; Search time 238 Seconds  
(without alignments)  
141.567 Million cell updates/sec

Title: US-10-062-624-40  
Perfect score: 1496  
Sequence: 1 MNKKLFTIINTVLVCLLSL.....AVATLNIYGGIGARLTF 293

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 440863 seqs, 114992915 residues

Total number of hits satisfying chosen parameters: 440863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pdb.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pdb.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pdb.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pdb.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pdb.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pdb.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pdb.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pdb.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pdb.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pdb.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pdb.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pdb.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pdb.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pdb.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1496	100.0	293	9	US-10-062-051-40
2	1496	100.0	293	9	US-10-062-920-40
3	1496	100.0	293	12	US-10-062-624-40
4	1483	99.1	293	9	US-10-314-639-44
5	1483	99.1	293	12	US-10-059-964-44
6	1072	71.7	291	9	US-10-314-639-18
7	1072	71.7	291	12	US-10-059-964-18
8	1033	69.1	291	9	US-10-284-986-10
9	1033	69.1	291	10	US-09-846-808-10
10	1032.5	69.0	300	9	US-10-284-986-12
11	1032.5	69.0	300	10	US-09-846-808-12
12	1026	68.6	293	9	US-10-284-986-13
13	1026	68.6	293	10	US-09-846-808-13
14	1026	68.6	297	9	US-10-314-639-14
15	1026	68.6	297	12	US-10-059-964-14
16	1008.5	67.4	296	9	US-10-314-639-56
17	1008.5	67.4	296	12	US-10-059-964-56
18	1006.5	67.3	298	9	US-10-284-986-11
19	1006.5	67.3	298	9	US-10-314-639-52

20	1006.5	67.3	298	10	US-09-846-808-11	Sequence 11, Appl
21	1006.5	67.3	298	12	US-10-059-964-52	Sequence 52, Appl
22	980.5	65.5	300	9	US-10-314-639-50	Sequence 50, Appl
23	980.5	65.5	300	12	US-10-059-964-50	Sequence 50, Appl
24	918	61.4	299	9	US-10-314-639-46	Sequence 46, Appl
25	918	61.4	299	12	US-10-059-964-46	Sequence 46, Appl
26	789.5	52.8	279	9	US-10-314-639-60	Sequence 60, Appl
27	789.5	52.8	279	12	US-10-059-964-60	Sequence 60, Appl
28	789	52.7	285	9	US-10-284-986-9	Sequence 9, Appl
29	789	52.7	285	10	US-09-846-808-9	Sequence 9, Appl
30	786.5	52.6	279	9	US-10-314-639-24	Sequence 24, Appl
31	786.5	52.6	279	12	US-10-059-964-24	Sequence 24, Appl
32	784	52.4	285	9	US-10-314-639-30	Sequence 30, Appl
33	784	52.4	285	12	US-10-059-964-30	Sequence 30, Appl
34	779.5	52.1	279	9	US-10-284-986-6	Sequence 6, Appl
35	779.5	52.1	279	10	US-09-846-808-6	Sequence 6, Appl
36	768	51.3	283	9	US-10-284-986-7	Sequence 7, Appl
37	768	51.3	283	9	US-10-314-639-26	Sequence 26, Appl
38	768	51.3	283	10	US-09-846-808-7	Sequence 7, Appl
39	768	51.3	283	12	US-10-059-964-26	Sequence 26, Appl
40	740	49.5	281	9	US-10-314-639-58	Sequence 58, Appl
41	740	49.5	281	12	US-10-059-964-58	Sequence 58, Appl
42	737.5	49.3	275	9	US-10-284-986-8	Sequence 8, Appl
43	737.5	49.3	275	9	US-10-314-639-28	Sequence 28, Appl
44	737.5	49.3	275	10	US-09-846-808-8	Sequence 8, Appl
45	737.5	49.3	275	12	US-10-059-964-28	Sequence 28, Appl

## ALIGNMENTS

### RESULT 1

US-10-062-051-40  
Sequence 40, Application US/10062051  
Publication No. US20030073095A1  
GENERAL INFORMATION: David H.  
APPLICANT: Walker, David H.  
APPLICANT: McBride, Jere W.  
TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof  
FILE REFERENCE: D6152C1P2  
CURRENT APPLICATION NUMBER: US/10/062,051  
PRIOR FILING DATE: 2002-01-31  
PRIOR APPLICATION NUMBER: US/09/660,587  
PRIOR FILING DATE: 2000-09-12  
PRIOR APPLICATION NUMBER: 09/261,358  
PRIOR FILING DATE: 1999-03-03  
NUMBER OF SEQ ID NOS: 46  
SEQ ID NO 40  
LENGTH: 293  
TYPE: PRT  
ORGANISM: Ehrlichia canis  
FEATURE:  
OTHER INFORMATION: amino acid sequence of E. canis p28-1 protein  
US-10-062-051-40

Query Match	100.0%	Score 1496	DB 9	Length 293
Best Local Similarity	100.0%	Pred. No. 2.5e-127	Indels 0	Gaps 0
Matches 293	Conservative 0	Mismatches 0		
Qy	1	MNNKLFPTIINTVLVCLLSLPLNISSKAIINNNAKYYGLYISGOYKPSVSVFNSFKET 60		
Db	1	MNNKLFPTIINTVLVCLLSLPLNISSKAIINNNAKYYGLYISGOYKPSVSVFNSFKET 60		
Qy	61	NVITKLIALKVDSDIETKTDAVGSINSPNFTIPYTAVFQDQNSVNFNGTIGYTFAGT 120		
Db	61	NVITKLIALKVDSDIETKTDAVGSINSPNFTIPYTAVFQDQNSVNFNGTIGYTFAGT 120		
Qy	121	RVIEGSEYEFQVKNPGGYLTSDAYRYFALAREMKNSTPKKVSNSIFHTVMRNDGLS 180		
Db	121	RVIEGSEYEFQVKNPGGYLTSDAYRYFALAREMKNSTPKKVSNSIFHTVMRNDGLS 180		

QY 181 IISVIVNVCYDFSLNNLSISPIYCGAGVDAIEFFDVLHIKFAVQSKGIAYSLPSNISL 240  
DB 181 IISVIVNVCYDFSLNNLSISPIYCGAGVDAIEFFDVLHIKFAVQSKGIAYSLPSNISL 240  
QY 241 FASLYYHKVMGNQFNKLVNQHVAELASIPKITSATVATLNIGYFGGEIGARLTF 293  
DB 241 FASLYYHKVMGNQFNKLVNQHVAELASIPKITSATVATLNIGYFGGEIGARLTF 293

## RESULT 2

US-10-062-920-40  
; Sequence 40, Application US/10062920  
; Publication No. US20030096250A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: US/10/062,920  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: US/09/660,587  
; PRIOR FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 40  
; LENGTH: 293  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of E. canis p28-1 protein  
US-10-062-920-40

Query Match 100.0%; Score 1496; DB 9; Length 293;  
Best Local Similarity 100.0%; Pred. No. 2.5e-127; Indels 0; Gaps 0;  
Matches 293; Conservative 0; Mismatches 0;  
QY 1 MNKLKFTIINTVLVCLLSLPNLSISSKAINNNAKYYGLYISGQKPSVSVFNSFKET 60  
DB 1 MNKLKFTIINTVLVCLLSLPNLSISSKAINNNAKYYGLYISGQKPSVSVFNSFKET 60  
QY 61 NVITKNLALKKQVDSIETKTDAISVGSINPSNFTPIPTAVFQDQSVNFGTIGYTFASGT 120  
DB 61 NVITKNLALKKQVDSIETKTDAISVGSINPSNFTPIPTAVFQDQSVNFGTIGYTFASGT 120  
QY 121 RVEIEGSVEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGLS 180  
DB 121 RVEIEGSVEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGLS 180  
QY 181 IISVIVNVCYDFSLNNLSISPIYCGAGVDAIEFFDVLHIKFAVQSKGIAYSLPSNISL 240  
DB 181 IISVIVNVCYDFSLNNLSISPIYCGAGVDAIEFFDVLHIKFAVQSKGIAYSLPSNISL 240  
QY 241 FASLYYHKVMGNQFNKLVNQHVAELASIPKITSATVATLNIGYFGGEIGARLTF 293  
DB 241 FASLYYHKVMGNQFNKLVNQHVAELASIPKITSATVATLNIGYFGGEIGARLTF 293

## RESULT 3

US-10-062-624-40  
; Sequence 40, Application US/10062624  
; Patent No. US20020115840A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-Kilodalton Immunodominant Protein  
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof  
; FILE REFERENCE: D6152CIP2/01  
; CURRENT APPLICATION NUMBER: US/10/062,624  
; CURRENT FILING DATE: 2002-01-31

; PRIOR APPLICATION NUMBER: 09/660,587  
; PRIOR FILING DATE: 2000-09-12  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 40  
; LENGTH: 293  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of E. canis p28-1 protein  
US-10-062-624-40

Query Match 100.0%; Score 1496; DB 12; Length 293;  
Best Local Similarity 100.0%; Pred. No. 2.5e-127; Indels 0; Gaps 0;  
Matches 293; Conservative 0; Mismatches 0;  
QY 1 MNKLKFTIINTVLVCLLSLPNLSISSKAINNNAKYYGLYISGQKPSVSVFNSFKET 60  
DB 1 MNKLKFTIINTVLVCLLSLPNLSISSKAINNNAKYYGLYISGQKPSVSVFNSFKET 60  
QY 61 NVITKNLALKKQVDSIETKTDAISVGSINPSNFTPIPTAVFQDQSVNFGTIGYTFASGT 120  
DB 61 NVITKNLALKKQVDSIETKTDAISVGSINPSNFTPIPTAVFQDQSVNFGTIGYTFASGT 120  
QY 121 RVEIEGSVEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGLS 180  
DB 121 RVEIEGSVEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGLS 180  
QY 181 IISVIVNVCYDFSLNNLSISPIYCGAGVDAIEFFDVLHIKFAVQSKGIAYSLPSNISL 240  
DB 181 IISVIVNVCYDFSLNNLSISPIYCGAGVDAIEFFDVLHIKFAVQSKGIAYSLPSNISL 240  
QY 241 FASLYYHKVMGNQFNKLVNQHVAELASIPKITSATVATLNIGYFGGEIGARLTF 293  
DB 241 FASLYYHKVMGNQFNKLVNQHVAELASIPKITSATVATLNIGYFGGEIGARLTF 293

## RESULT 4

US-10-314-639-44  
; Sequence 44, Application US/10314639  
; Publication No. US20030103991A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20030103991A1  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; TITLE OF INVENTION: Chaffensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/314,639  
; CURRENT FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US/09/314,701  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 44  
; LENGTH: 293  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
US-10-314-639-44

Query Match 99.1%; Score 1483; DB 9; Length 293;  
Best Local Similarity 99.3%; Pred. No. 3.7e-126; Indels 0; Gaps 0;  
Matches 291; Conservative 0; Mismatches 2;  
QY 1 MNKLKFTIINTVLVCLLSLPNLSISSKAINNNAKYYGLYISGQKPSVSVFNSFKET 60  
DB 1 MNKLKFTIINTVLVCLLSLPNLSISSKAINNNAKYYGLYISGQKPSVSVFNSFKET 60  
QY 61 NVITKNLALKKQVDSIETKTDAISVGSINPSNFTPIPTAVFQDQSVNFGTIGYTFASGT 120  
DB 61 NVITKNLALKKQVDSIETKTDAISVGSINPSNFTPIPTAVFQDQSVNFGTIGYTFASGT 120  
QY 121 RVEIEGSVEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGLS 180  
DB 121 RVEIEGSVEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGLS 180



Qy 181 IISVIVNVCYDFSLNNLSISPIYICGGAGVDAIEFDFVLHIKFAVQSKLGIAYSPLPSNISL 240  
 Db 181 IISVIVNVCYDFSLNNLSISPIYICGGAGVDAIEFDFVLHIKFAVQSKLGIAYSPLPSNISL 240  
 Qy 241 FASLYYHKVMGNQFNKLNVOHVAELASIPKITSATVATLNIGYFGGEIGARLTF 293  
 Db 241 FASLYYHKVMGNQFNKLNVOHVAELASIPKITSATVATLNIGYFGGEIGARLTF 293

RESULT 5  
 US-10-059-964-44  
 ; Sequence 44, Application US/10059964  
 ; Patent No. US20020120115A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rikihisa, Yasuko  
 ; APPLICANT: Ohasi, No. US20020120115A1  
 ; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
 ; TITLE OF INVENTION: Chaffeensis  
 ; FILE REFERENCE: 22727/04021  
 ; CURRENT APPLICATION NUMBER: US/10/059,964  
 ; CURRENT FILING DATE: 2002-01-28  
 ; EARLIER APPLICATION NUMBER: 09/314,701  
 ; EARLIER FILING DATE: 1999-05-19  
 ; NUMBER OF SEQ ID NOS: 66  
 ; SOFTWARE: Patent In Ver. 2.0  
 ; SEQ ID NO 44  
 ; LENGTH: 293  
 ; TYPE: PRT  
 ; ORGANISM: Ehrlichia canis  
 US-10-059-964-44

Query Match 99.1%; Score 1483; DB 12; Length 293;  
 Best Local Similarity 99.3%; Pred. No. 3.7e-126;  
 Matches 291; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MNKKLFTIINTVLVCLLSLPISSKAINNNKAYGLYISGOYKPSVSFVSFVKET 60  
 Db 1 MNKKLFTIINTVLVCLLSLPISSKAINNNKAYGLYISGOYKPSVSFVSFVKET 60  
 Qy 61 NVITKNLIALKQVDSTETKTDAISVIGSNPSNFTIPYTAVFQDNSVNFNGTIGYTFAGT 120  
 Db 61 NVITKNLIALKQVDSTETKTDAISVIGSNPSNFTIPYTAVFQDNSVNFNGTIGYTFAGT 120  
 Qy 121 RVEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGLS 180  
 Db 121 RVEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGLS 180  
 Qy 181 IISVIVNVCYDFSLNNLSISPIYICGGAGVDAIEFDFVLHIKFAVQSKLGIAYSPLPSNISL 240  
 Db 181 IISVIVNVCYDFSLNNLSISPIYICGGAGVDAIEFDFVLHIKFAVQSKLGIAYSPLPSNISL 240  
 Qy 241 FASLYYHKVMGNQFNKLNVOHVAELASIPKITSATVATLNIGYFGGEIGARLTF 293  
 Db 241 FASLYYHKVMGNQFNKLNVOHVAELASIPKITSATVATLNIGYFGGEIGARLTF 293

RESULT 6  
 US-10-314-639-18  
 ; Sequence 18, Application US/10314639  
 ; Publication No. US20030103991A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rikihisa, Yasuko  
 ; APPLICANT: Ohasi, No. US20030103991A1  
 ; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
 ; TITLE OF INVENTION: Chaffeensis  
 ; FILE REFERENCE: 22727/04021  
 ; CURRENT APPLICATION NUMBER: US/10/314,639  
 ; CURRENT FILING DATE: 2002-12-09  
 ; EARLIER APPLICATION NUMBER: US/09/314,701  
 ; EARLIER FILING DATE: 1999-05-19  
 ; NUMBER OF SEQ ID NOS: 66  
 ; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 18  
 ; LENGTH: 291  
 ; TYPE: PRT  
 ; ORGANISM: Ehrlichia chaffeensis  
 US-10-314-639-18

Query Match 71.7%; Score 1072; DB 9; Length 291;  
 Best Local Similarity 70.3%; Pred. No. 5.3e-89;  
 Matches 206; Conservative 33; Mismatches 52; Indels 2; Gaps 2;

Qy 1 MNKKLFTIINTVLVCLLSLPISSKAINNNKAYGLYISGOYKPSVSFVSFVKET 60  
 Db 1 MNKKLFTIINTVLVCLLSLPISSKAINNNKAYGLYISGOYKPSVSFVSFVKET 59  
 Qy 61 NVITKNLIALKQVDSTETKTDAISVIGSNPSNFTIPYTAVFQDNSVNFNGTIGYTFAGT 120  
 Db 60 NTITKNLIALKQDINSLEVNADASQGISHPGNFTIPYIAAFEDNAFNGAIGY-ITEGL 118  
 Qy 121 RVEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGLS 180  
 Db 119 RIEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGLS 178  
 Qy 181 IISVIVNVCYDFSLNNLSISPIYICGGAGVDAIEFDFVLHIKFAVQSKLGIAYSPLPSNISL 240  
 Db 179 IISIMVNGCYDFSSDNLVSPYICGGIGVDIAIEFFDHALHIKLAQCQSKLGITYQLSYNISL 238  
 Qy 241 FASLYYHKVMGNQFNKLNVOHVAELASIPKITSATVATLNIGYFGGEIGARLTF 293  
 Db 239 FADGYHQVIGNQFNKLNVOHVAELADPAKTSATVATLNIGYFGAEVGVRFIP 291

## RESULT 7

US-10-059-964-18  
 ; Sequence 18, Application US/10059964  
 ; Patent No. US20020120115A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rikihisa, Yasuko  
 ; APPLICANT: Ohasi, No. US20020120115A1  
 ; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
 ; TITLE OF INVENTION: Chaffeensis  
 ; FILE REFERENCE: 22727/04021  
 ; CURRENT APPLICATION NUMBER: US/10/059,964  
 ; CURRENT FILING DATE: 2002-01-28  
 ; EARLIER APPLICATION NUMBER: 09/314,701  
 ; EARLIER FILING DATE: 1999-05-19  
 ; NUMBER OF SEQ ID NOS: 66  
 ; SOFTWARE: Patent In Ver. 2.0  
 ; SEQ ID NO 18  
 ; LENGTH: 291  
 ; TYPE: PRT  
 ; ORGANISM: Ehrlichia chaffeensis  
 US-10-059-964-18

Query Match 71.7%; Score 1072; DB 12; Length 291;  
 Best Local Similarity 70.3%; Pred. No. 5.3e-89;  
 Matches 206; Conservative 33; Mismatches 52; Indels 2; Gaps 2;

Qy 1 MNKKLFTIINTVLVCLLSLPISSKAINNNKAYGLYISGOYKPSVSFVSFVKET 60  
 Db 1 MNKKLFTIINTVLVCLLSLPISSKAINNNKAYGLYISGOYKPSVSFVSFVKET 59  
 Qy 61 NVITKNLIALKQVDSTETKTDAISVIGSNPSNFTIPYTAVFQDNSVNFNGTIGYTFAGT 120  
 Db 60 NTITKNLIALKQDINSLEVNADASQGISHPGNFTIPYIAAFEDNAFNGAIGY-ITEGL 118  
 Qy 121 RVEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGLS 180  
 Db 119 RIEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGLS 178  
 Qy 181 IISVIVNVCYDFSLNNLSISPIYICGGAGVDAIEFDFVLHIKFAVQSKLGIAYSPLPSNISL 240  
 Db 179 IISIMVNGCYDFSSDNLVSPYICGGIGVDIAIEFFDHALHIKLAQCQSKLGITYQLSYNISL 238



QY 234 LPSNLSFASLYHKVMGNQKLNQVHVAELASIPKITSAVATLNIGYFGGEIGARLTF 293  
DB 241 LSSNVSLFAGGYHQVMGNQKLNQVHVAELNDAPKVTSAVATLDIGYFGGEIGARLIF 300

## RESULT 11

US-09-846-808-12  
; Sequence 12, Application US/09846808  
; Patent No. US20020064531A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: Yu, Xu-Jie  
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane  
; TITLE OF INVENTION: Protein Multigene Family  
; FILE REFERENCE: D6311  
; CURRENT APPLICATION NUMBER: US/09/846.808  
; CURRENT FILING DATE: 2001-05-01  
; PRIOR APPLICATION NUMBER: 60/201,035  
; PRIOR FILING DATE: 2000-05-01  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 12  
; LENGTH: 300  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: P28-12 Outer Membrane Protein of  
; OTHER INFORMATION: Ehrlichia chaffeensis  
US-09-846-808-12

Query Match 69.0%; Score 1032.5; DB 10; Length 300;  
Best Local Similarity 66.7%; Pred. No. 2.1e-85;  
Matches 200; Conservative 35; Mismatches 58; Indels 7; Gaps 2;  
QY 1 MNKLKFTIINTVLVCLLSLPSNISKAINNNKAYGLYISGOYKPSVSVFSNFSVKET 60  
DB 1 MKKKNQFTITSTILVCLLSLNSASLNTNSSTKKQGLYSGQYKPSVSVFSNFSVKET 60  
QY 61 NVITKNLIALKQVDSIETKTDAVGISNPSNFTIPYTAVFQDQNSVNFNGTIGYFAEGT 120  
DB 61 NFPTKYLAALKKDINSVEFSDSVTAGISYPLNFSTPYIAVFQDQNSVNFNGTIGYFAEGT 120  
QY 121 RVEIEGSVEEFDVKNPGGYT-LSDAYRYFALAREMKNSTFPKPKVSN-----SIFHTV 173  
DB 121 RIEIEGPHERFDVKNPGGYT-LSDAYRYFALAREMKNSTFPKPKVSN-----SIFHTV 173  
QY 174 MRNDGLSIIVNVVVDVDFSLNLSISPYICGGAGVDAIEFFDLVHLIKFAYQSKGLIAYS 233  
DB 181 MKNEGLSIIVNVVVDVDFSLNLSISPYICGGAGVDAIEFFDLVHLIKFAYQSKGLIAYS 233  
QY 234 LPSNLSFASLYHKVMGNQKLNQVHVAELASIPKITSAVATLNIGYFGGEIGARLTF 293  
DB 241 LSSNVSLFAGGYHQVMGNQKLNQVHVAELNDAPKVTSAVATLDIGYFGGEIGARLIF 300

## RESULT 12

US-10-284-986-13  
; Sequence 13, Application US/10284986  
; Publication No. US20030091588A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: Yu, Xu-Jie  
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane  
; TITLE OF INVENTION: Protein Multigene Family  
; FILE REFERENCE: D6311D1  
; CURRENT APPLICATION NUMBER: US/10/284.986  
; CURRENT FILING DATE: 2002-10-31  
; PRIOR APPLICATION NUMBER: 09/846,808  
; PRIOR FILING DATE: 2001-05-01  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 13  
; LENGTH: 293  
; TYPE: PRT

; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: P28-13 Outer Membrane Protein of  
; OTHER INFORMATION: Ehrlichia chaffeensis  
US-10-284-986-13

Query Match 68.6%; Score 1026; DB 9; Length 293;  
Best Local Similarity 66.0%; Pred. No. 7.8e-85;  
Matches 194; Conservative 40; Mismatches 58; Indels 2; Gaps 2;  
QY 1 MNKLKFTIINTVLVCLLSLPSNISKAINNNKAYGLYISGOYKPSVSVFSNFSVKET 60  
DB 1 MKKKNQFTITSTAMVCLLLPGISFSSTINNSAKQGLYISGOYKPSVSVFSNFSVKET 60  
QY 61 NVITKNLIALKQVDSIETKTDAVGISNPSNFTIPYTAVFQDQNSVNFNGTIGYFAEGT 120  
DB 61 NVPTKQLIALKQDINSVAVGNSNATTGISNPGNFTIPYTAEFQDNVANFNGAVGYSFFDSL 120  
QY 121 RVEIEGSVEEFDVKNPGGYT-LSDAYRYFALAREMKNSTFPKPKVSN-----SIFHTV 179  
DB 121 RIEIEGPHERFDVKNPGGYTQVKDAYRYFALARDLKGDFPEPKAE-DTGVVHTVMKNDGL 179  
QY 180 SIISVIVNVVDVDFSLNLSISPYICGGAGVDAIEFFDLVHLIKFAYQSKGLIAYS 239  
DB 180 SIISTMVNVVDVDFSLNLSISPYICAGMNAIEFFDLVHLIKFAYQSKGLIAYS 239  
QY 240 LFASLYHKVMGNQKLNQVHVAELASIPKITSAVATLNIGYFGGEIGARLTF 293  
DB 240 LFLDGYHQVIGNQKLNQVHVAELASIPKITSAVATLNIGYFGGEIGARLTF 293

## RESULT 13

US-09-846-808-13  
; Sequence 13, Application US/09846808  
; Patent No. US20020064531A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: Yu, Xu-Jie  
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane  
; TITLE OF INVENTION: Protein Multigene Family  
; FILE REFERENCE: D6311  
; CURRENT APPLICATION NUMBER: US/09/846.808  
; CURRENT FILING DATE: 2001-05-01  
; PRIOR APPLICATION NUMBER: 60/201,035  
; PRIOR FILING DATE: 2000-05-01  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 13  
; LENGTH: 293  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: P28-13 Outer Membrane Protein of  
; OTHER INFORMATION: Ehrlichia chaffeensis  
US-09-846-808-13

Query Match 68.6%; Score 1026; DB 10; Length 293;  
Best Local Similarity 66.0%; Pred. No. 7.8e-85;  
Matches 194; Conservative 40; Mismatches 58; Indels 2; Gaps 2;  
QY 1 MNKLKFTIINTVLVCLLSLPSNISKAINNNKAYGLYISGOYKPSVSVFSNFSVKET 60  
DB 1 MKKKNQFTITSTAMVCLLLPGISFSSTINNSAKQGLYISGOYKPSVSVFSNFSVKET 60  
QY 61 NVITKNLIALKQVDSIETKTDAVGISNPSNFTIPYTAVFQDQNSVNFNGTIGYFAEGT 120  
DB 61 NVPTKQLIALKQDINSVAVGNSNATTGISNPGNFTIPYTAEFQDNVANFNGAVGYSFFDSL 120  
QY 121 RVEIEGSVEEFDVKNPGGYT-LSDAYRYFALAREMKNSTFPKPKVSN-----SIFHTV 179  
DB 121 RIEIEGPHERFDVKNPGGYTQVKDAYRYFALARDLKGDFPEPKAE-DTGVVHTVMKNDGL 179  
QY 180 SIISVIVNVVDVDFSLNLSISPYICGGAGVDAIEFFDLVHLIKFAYQSKGLIAYS 239  
DB 180 SIISTMVNVVDVDFSLNLSISPYICAGMNAIEFFDLVHLIKFAYQSKGLIAYS 239

Db 180 SILSTMVNCYDFSVDELPLVPYICAGMGINAIEFFDALHVKFAYQKLGISYQLFTKVN 239  
QY 240 LFASLYYHKVMGNQFKNLNVQHVLAELASIPKITSATVATLNIGYFGEIGARLTF 293  
Db 240 LFLDGYHQVIGNQFKNLNVNHVYTLKESPKVTSATVATLDIAFYGGEVGIIRTF 293

## RESULT 14

US-10-314-639-14  
; Sequence 14, Application US/10314639  
; Publication No. US20030103991A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20030103991A1io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; TITLE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/314,639  
; CURRENT FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US/09/314,701  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 297  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
US-10-314-639-14

Query Match 68.6%; Score 1026; DB 9; Length 297;  
Best Local Similarity 66.0%; Pred. No. 7.9e-85;  
Matches 194; Conservative 40; Mismatches 58; Indels 2; Gaps 2;  
QY 1 MNKKLKFITINTVLVCLLSLNPNISSKAINNNAKYGLISGQKPSVSVFSPSVK 60  
Db 5 MNKKNKFTTISTAMVCLLLPGISFSETINNSAKQPGLYISGQKPSVSVFSPSVK 64  
QY 61 NVITKNLIALKQDVDSIETKTDAISGINSNPSNFTPIPTAVFQDINSVNFNGTIGYTF 120  
Db 65 NVPTKQLIALKDDINSVAVGNSNATTGINSNPGNFTPIPTAEFQDNVANFNGAVGYSFPD 124  
QY 121 RVEIEGSVEEFVKNPGGYT-LSDAYRYFALAREMKNSTPPEKVSNSIFHTVMRNDGL 179  
Db 125 RIEIEGFHEKFDVKNPGGYTQVKDAYRYFALARDLKDGFEPKAE-DTGVTYHTVMKNDGL 183  
QY 180 SIISVIVNVCYDFSLNNLSISPYICGAGVDAIEFFDVLHIKFAYQSKLGIAYSLP 239  
Db 184 SILSTMVNCYDFSVDELPLVPYICAGMGINAIEFFDALHVKFAYQKLGISYQLFTKVN 243  
QY 240 LFASLYYHKVMGNQFKNLNVQHVLAELASIPKITSATVATLNIGYFGEIGARLTF 293  
Db 244 LFLDGYHQVIGNQFKNLNVNHVYTLKESPKVTSATVATLDIAFYGGEVGIIRTF 297

## RESULT 15

US-10-059-964-14  
; Sequence 14, Application US/10059964  
; Patent No. US20020120115A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20020120115A1io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; TITLE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/059,964  
; CURRENT FILING DATE: 2002-01-28  
; EARLIER APPLICATION NUMBER: 09/314,701  
; EARLIER FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 297  
; TYPE: PRT

; ORGANISM: Ehrlichia chaffeensis  
US-10-059-964-14  
Query Match 68.6%; Score 1026; DB 12; Length 297;  
Best Local Similarity 66.0%; Pred. No. 7.9e-85;  
Matches 194; Conservative 40; Mismatches 58; Indels 2; Gaps 2;  
QY 1 MNKKLKFITINTVLVCLLSLNPNISSKAINNNAKYGLISGQKPSVSVFSPSVK 60  
Db 5 MNKKNKFTTISTAMVCLLLPGISFSETINNSAKQPGLYISGQKPSVSVFSPSVK 64  
QY 61 NVITKNLIALKQDVDSIETKTDAISGINSNPSNFTPIPTAVFQDINSVNFNGTIGYTF 120  
Db 65 NVPTKQLIALKDDINSVAVGNSNATTGINSNPGNFTPIPTAEFQDNVANFNGAVGYSFPD 124  
QY 121 RVEIEGSVEEFVKNPGGYT-LSDAYRYFALAREMKNSTPPEKVSNSIFHTVMRNDGL 179  
Db 125 RIEIEGFHEKFDVKNPGGYTQVKDAYRYFALARDLKDGFEPKAE-DTGVTYHTVMKNDGL 183  
QY 180 SIISVIVNVCYDFSLNNLSISPYICGAGVDAIEFFDVLHIKFAYQSKLGIAYSLP 239  
Db 184 SILSTMVNCYDFSVDELPLVPYICAGMGINAIEFFDALHVKFAYQKLGISYQLFTKVN 243  
QY 240 LFASLYYHKVMGNQFKNLNVQHVLAELASIPKITSATVATLNIGYFGEIGARLTF 293  
Db 244 LFLDGYHQVIGNQFKNLNVNHVYTLKESPKVTSATVATLDIAFYGGEVGIIRTF 297

## RESULT 16

US-10-314-639-56  
; Sequence 56, Application US/10314639  
; Publication No. US20030103991A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20030103991A1io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; TITLE OF INVENTION: Chaffeensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/314,639  
; CURRENT FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US/09/314,701  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 56  
; LENGTH: 296  
; TYPE: PRT  
; ORGANISM: p30-7  
US-10-314-639-56

Query Match 67.4%; Score 1008.5; DB 9; Length 296;  
Best Local Similarity 67.8%; Pred. No. 3e-83;  
Matches 200; Conservative 36; Mismatches 54; Indels 5; Gaps 4;  
QY 1 MNKKLKFITINTVLVCLLSLNPNISSKAINNNAKYGLISGQKPSVSVFSPSVK 60  
Db 5 MNKKNKFLIRIFLTCMLSLNPNISSLSK-VNN--BKHSGLYISGQKPSVSVFSPSVK 61  
QY 61 NVITKNLIALKQDVDSIETKTDA-SVGINSNPSNFTPIPTAVFQDINSVNFNGTIGYTF 119  
Db 62 NFHTKHLIALKQDVDSVEIDTGSNTAGISNPSNFTPIPTAEFQDNHTNCNGSIGVAF 121  
QY 120 TRVIEGSVEEFVKNPGGYT-LSDAYRYFALAREMKNSTPPEKVSNSIFHTVMRNDG 178  
Db 122 PRIEIELSYEKFDVKNPGGYTQVKDAYRYFALAREINISLFPQKQKESGIYHVVMKNDG 181  
QY 179 LSISVIVNVCYDFSLNNLSISPYICGAGVDAIEFFDVLHIKFAYQSKLGIAYSLP 238  
Db 182 LSILSNVNICYDFSLNNLPISPYICGAGMGNAIEFFDALHVKFAYQSKAGISYQLLRKI 241  
QY 239 SLFASLYYHKVMGNQFKNLNVQHVLAELASIPKITSATVATLNIGYFGEIGARLTF 293  
Db 242 NLFIDVYVYEVISNKFKNLKVQHVHVKDNPKVTSATVATLDIAFYGSEAGIRIIF 296

```

RESULT 17
US-10-059-964-56
; Sequence 56, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohashi, No. US20020120115A1io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; CURRENT FILING DATE: 2002-01-28
; EARLIER APPLICATION NUMBER: 09/314,701
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 56
; LENGTH: 296
; TYPE: PRT
; ORGANISM: p30-7
US-10-059-964-56

Query Match          67.4%; Score 1008.5; DB 12; Length 296;
Best Local Similarity 67.8%; Pred. No. 3e-83;
Matches 200; Conservative 36; Mismatches 54; Indels 5; Gaps 4

Qy      1  MNKKLKEITINTVLVCLLSLNLSSKAINNAKKYGLYISGQVKPSVSPFSNFSVKET 60
Db      5  MNKKSQFLIRIFITCMLSLNLNLSSK--VNN--EKHSGLYISGQVKPSVSPFSNFSVKET 61

Qy      61 NVITPKNLIALKQVDSDISBTKTDA--SVGISNPSNFTIPYTAVFQDNVSNFNGTIGYTFAG 119
Db      62 NFHTKHLIALKQVDSDVEIDTGSNTAGISNPSNFTIPYTAERQDNHTNCNGSIGYAFAG 121

Qy      120 TRVEIGSYEEFDVKNPGCY--TLSDAYRYFALAREMKNSTFPKPKVNSIFHTVMRNDG 178
Db      122 PRBIEIUSYEKFDVKNPTGYTTVKDAYRYFALAREINSLFPQKQKSGSIGYHVMKNDG 181

Qy      179 LSIISVTVNVCYDFPSLNLNLSIPYICGGAGVDATFEFFDLVHLKPAQSKGLIAYSILPSNI 238
Db      182 LSILSNVNICDFPSLNLNLPISPLYCGMGINAEFFDLAHVKFAYQSKAGISYQLLRKI 241

Qy      239 SLFASLYYHKVGNQFKNLVQHVABLASIPKITSAVATLNIGYFGGIGARLTF 293
Db      242 NLFIDVYYEIVSNKFKLQVHVHELKDNPKVTSAVATLDTIAYFGSGAGIRIIF 296

```

```

Query Match          67.3%; Score 1006.5; DB 9; Length 298;
Best Local Similarity 65.2%; Pred. No. 4.6e-83;
Matches 193; Conservative 40; Mismatches 56; Indels 7; Gaps 3;

QY      1 MNKKLKFTIINTVLVCLLSLPNITSSSKAINNAKKYVGLYISQYKPSVVFSGNFVKET 60
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      1 MNHKSMLFTTGTALISLLSPNVFSFGIINNANN-LGIYISQYKPSVVFSGNFVKET 59
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY     61 NVITKNIJALKQDVDSIETKTDASVGSINPSNFTPIPYTAVFQDNSVNFNGTIGYTFAGT 120
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     60 NFFTQQLVALKKDIDSVIDSTNADSGINNPNQFTPIYPKPFQDNAASFSGALGFFYARGL 119
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY    121 RVEIEGSYEEDVKNGPGYT-LSDARYYFALAREMKGNSTFPKEKVS-----NSIFHTVM 174
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db    120 RLEMEGSYEEDVKNGPGYTKVKDARYYFALAREMQSGQTCPKHKETSGIQPHGIYHTVM 179
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY    175 RNDGLSITISVINVCYDFSLNNLSISPIYICGGAGVDAIBFFDVLHIKFAYOSKLGIAYSL 234
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db    180 RNDGVSISSVIINGCYNFTLSNIPSPYMCVGVGIDAIQFDSLHIKFHQSKLGIITYPL 239
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY    235 PSNISLFSALYYHKVMGNQFKNLNVQHVAEIASIPKITSAVATINIGYFQGEIGAR 290
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db    240 SSNVHLFADSYHHKVGINKFKNLRVQHVELQQVPKVTSAVATLIDIGYFGGEGVGR 295
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 19
US-10-314-639-52
; Sequence 52, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20030103991A1io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/314,701
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 52
; LENGTH: 298
; TYPE: PRT
; ORGANISM: OMP-1H
US-10-314-639-52

```

## RESULT 20

```

US-09-846-808-11
; Sequence 11, Application US/09846808
; Patent No. US20020064531A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xu-Jie
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane
; TITLE OF INVENTION: Protein Multigene Family
; FILE REFERENCE: D6311
; CURRENT APPLICATION NUMBER: US/09/846,808
; CURRENT FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 60/201,035
; PRIOR FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 11
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: P28-11 Outer Membrane Protein of
; OTHER INFORMATION: Ehrlichia chaffeensis
US-09-846-808-11

Query Match 67.3%; Score 1006.5; DB 10; Length 298;
Best Local Similarity 65.2%; Pred. No. 4.6e-83;
Matches 193; Conservative 40; Mismatches 56; Indels 7; Gaps 3;

QY 1 MNNKLTFTINTVLVCLLSLNPNISSKAINNNAKYYGLYISGQYKPSVFSNFSVKET 60
DB 1 MNNKSMFTTIGTALISLLSLNPVSGIINNANN-LGIYISGQYKPSVFSNFSVKET 59
;
QY 61 NVITKNLIALKKVDISTETKTDAVGSINPSNFTPIPTAVFQDQNSVNFNGTIGYTFASGT 120
DB 60 NFTTQOLVALKKDIDSDVISTNADSGINNPNQFTPIPIPKFQDQNAASFGALGFYARGL 119
;
QY 121 RVEIEGSYEEDVKNPGGYT-LSDAYRYFALAREMKGNSFTPKKVS-----NSIFHTVM 174
DB 120 RLEMEGSYEEDVKNPGGYTKVDAYRYFALAREMQSGQTCPKHKETSGIQPHGIYHTVM 179
;
QY 175 RNDGLSIISVIVNVCYDFSLNNLSISPIYICGAGVDAIEFFDVLHIKFAYOSKLGIAVSL 234
DB 180 RNDGVSISVIINGCYNFTLSNLPISPYMCVGMGIDAIQFFDSLHIKFAHQSKLGITYPL 239
;
QY 235 PSNISLFLASYHYHKVMGNQFNLVQHVLAELASIPKITSAVATLNIGYFGGEIGAR 290
DB 240 SSVNHLFADSYHYHKVIGNKFNLRVQHVYELQQVPKVTSAVATLDIGYFGGEVGR 295
;
; ORGANISM: OMP-1H
US-10-059-964-52

Query Match 67.3%; Score 1006.5; DB 12; Length 298;
Best Local Similarity 65.2%; Pred. No. 4.6e-83;
Matches 193; Conservative 40; Mismatches 56; Indels 7; Gaps 3;

QY 1 MNNKLTFTINTVLVCLLSLNPNISSKAINNNAKYYGLYISGQYKPSVFSNFSVKET 60
DB 1 MNNKSMFTTIGTALISLLSLNPVSGIINNANN-LGIYISGQYKPSVFSNFSVKET 59
;
QY 61 NVITKNLIALKKVDISTETKTDAVGSINPSNFTPIPTAVFQDQNSVNFNGTIGYTFASGT 120
DB 60 NFTTQOLVALKKDIDSDVISTNADSGINNPNQFTPIPIPKFQDQNAASFGALGFYARGL 119
;
QY 121 RVEIEGSYEEDVKNPGGYT-LSDAYRYFALAREMKGNSFTPKKVS-----NSIFHTVM 174
DB 120 RLEMEGSYEEDVKNPGGYTKVDAYRYFALAREMQSGQTCPKHKETSGIQPHGIYHTVM 179
;
QY 175 RNDGLSIISVIVNVCYDFSLNNLSISPIYICGAGVDAIEFFDVLHIKFAYOSKLGIAVSL 234
DB 180 RNDGVSISVIINGCYNFTLSNLPISPYMCVGMGIDAIQFFDSLHIKFAHQSKLGITYPL 239
;
QY 235 PSNISLFLASYHYHKVMGNQFNLVQHVLAELASIPKITSAVATLNIGYFGGEIGAR 290
DB 240 SSVNHLFADSYHYHKVIGNKFNLRVQHVYELQQVPKVTSAVATLDIGYFGGEVGR 295
;
; ORGANISM: OMP-1H
US-10-059-964-52

Query Match 67.3%; Score 1006.5; DB 12; Length 298;
Best Local Similarity 65.2%; Pred. No. 4.6e-83;
Matches 193; Conservative 40; Mismatches 56; Indels 7; Gaps 3;

```

```

QY 1 MNNKLTFTINTVLVCLLSLNPNISSKAINNNAKYYGLYISGQYKPSVFSNFSVKET 60
DB 1 MNNKSMFTTIGTALISLLSLNPVSGIINNANN-LGIYISGQYKPSVFSNFSVKET 59
;
QY 61 NVITKNLIALKKVDISTETKTDAVGSINPSNFTPIPTAVFQDQNSVNFNGTIGYTFASGT 120
DB 60 NFTTQOLVALKKDIDSDVISTNADSGINNPNQFTPIPIPKFQDQNAASFGALGFYARGL 119
;
QY 121 RVEIEGSYEEDVKNPGGYT-LSDAYRYFALAREMKGNSFTPKKVS-----NSIFHTVM 174
DB 120 RLEMEGSYEEDVKNPGGYTKVDAYRYFALAREMQSGQTCPKHKETSGIQPHGIYHTVM 179
;
QY 175 RNDGLSIISVIVNVCYDFSLNNLSISPIYICGAGVDAIEFFDVLHIKFAYOSKLGIAVSL 234
DB 180 RNDGVSISVIINGCYNFTLSNLPISPYMCVGMGIDAIQFFDSLHIKFAHQSKLGITYPL 239
;
QY 235 PSNISLFLASYHYHKVMGNQFNLVQHVLAELASIPKITSAVATLNIGYFGGEIGAR 290
DB 240 SSVNHLFADSYHYHKVIGNKFNLRVQHVYELQQVPKVTSAVATLDIGYFGGEVGR 295
;
; ORGANISM: OMP-1H
US-10-314-639-50

Query Match 65.5%; Score 980.5; DB 9; Length 300;
Best Local Similarity 63.6%; Pred. No. 1e-80;
Matches 194; Conservative 35; Mismatches 59; Indels 17; Gaps 3;

```

```

RESULT 22
US-10-314-639-50
; Sequence 50, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20030103991A110
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/314,701
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 50
; LENGTH: 300
; TYPE: PRT
; ORGANISM: OMP-1Z
US-10-314-639-50

```

```

Query Match 65.5%; Score 980.5; DB 9; Length 300;
Best Local Similarity 63.6%; Pred. No. 1e-80;
Matches 194; Conservative 35; Mismatches 59; Indels 17; Gaps 3;

```

```

QY 1 MNNKLTFTINTVLVCLLSLNPNISSKAINNNAKYYGLYISGQYKPSVFSNFSVKET 60
DB 1 MNNKQFTITISTILVCLLSLSNASLNTNSSTKKQFGLVSGQYKPSVFSNFSVKET 60
;
QY 61 NVITKNLIALKKVDISTETKTDAVGSINPSNFTPIPTAVFQDQNSVNFNGTIGYTFASGT 120
DB 61 NFPTKYLAALKDIDINSVEFDDSVTAGISYPLNFTPIPIAVFQDQNSVNFNGAIGYTFVGGP 120
;
QY 121 RVEIEGSYEEDVKNPGGYTLDAYRY-----PALAREMKGNSFTPKKVS-----S 168
DB 121 RIEIEGSYEEDVKNP-----LEDIQYKMHMTVDPALARDIDISPTSKNRTSDGNSSYK 175
;
QY 169 IFHTVMRNDGLSIISVIVNVCYDFSLNNLSISPIYICGAGVDAIEFFDVLHIKFAYOSKL 228
DB 176 VYHTVMKNEGLSIISVIVNVCYDFSLNNLSILPVCVCGGIGVNALEFFDALHVKACQCKL 235
;
QY 229 GIAYSLPSNISLFLASYHYHKVMGNQFNLVQHVLAELASIPKITSAVATLNIGYFGGEIG 288
DB 236 GITYPSSNVSFLFAGGYHYHVMGNQFNLVQHVLAELNDAPKVTSAVATLDIGYFGGEIG 295
;
QY 289 ARLTF 293
DB 296 ARLIF 300

```

## RESULT 23

US-10-059-964-50  
; Sequence 50, Application US/10059964  
; Patent No. US20020120115A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20020120115A1io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/059,964  
; CURRENT FILING DATE: 2002-01-28  
; EARLIER APPLICATION NUMBER: 09/314,701  
; EARLIER FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 50  
; LENGTH: 300  
; TYPE: PRT  
; ORGANISM: OMP-12  
; US-10-059-964-50

Query Match 65.5%; Score 980.5; DB 12; Length 300;  
Best Local Similarity 63.6%; Pred. No. 1e-80;  
Matches 194; Conservative 35; Mismatches 59; Indels 17; Gaps 3;

Qy	1	MNNKLFPTIINTVLVCLLSLPNISSSKAINNNAKYYGLYISGOYKPSVSFNSFKVET	60
Db	1	MKKKQFPTITILVCLLSLNSLNTTSSTKQFGLYVSGQYKPSVIFNSFKVET	60
Qy	61	NVITKNIALKKQVDSIETKTDASVGISNPSNFTPIPTAVFQDQNSVNFNGTIGYTPAEGT	120
Db	61	NFTPKYLAALKKQINSVEFDDSVTAGISYPLNFSTPIYIAVFQDQNSVNFNGTIGYTPAEGT	120
Qy	121	RVEIEGSEYEEFDKQNGPGGYTLSDAYR-----PALAREMKGNSFTPEKVS-----S	168
Db	121	RRIEIEGSEYEEFDKQNGPGGYTLSDAYR-----LEDQYKMHVTVDPALARDISPTSPKRTSHDGNSSYK	175
Qy	169	IFHTWMNDGLSIISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDVLHIKPAYQSKL	228
Db	176	VYHTVMKNEGLSIISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDVLHIKPAYQSKL	235
Qy	229	GIATSLPSNLSLFAVYHVKMGQNFKNLVQHVLAELASIPKITSAVATLNIGYFGGEIG	289
Db	236	GITYPLSSNVSLFAGGYHQWGMQNFKNLVQHVLAELASIPKITSAVATLNIGYFGGEIG	295
Qy	289	ARLTF	293
Db	296	ARLIF	300

## RESULT 24

US-10-314-639-46  
; Sequence 46, Application US/10314639  
; Publication No. US20030103991A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20030103991A1io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/314,639  
; CURRENT FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US/09/314,701  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 46  
; LENGTH: 299  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
; US-10-314-639-46

Query Match 61.4%; Score 918; DB 9; Length 299;  
Best Local Similarity 60.5%; Pred. No. 4.7e-75;  
Matches 181; Conservative 41; Mismatches 69; Indels 8; Gaps 4;

Qy	1	MNNKLFPTIINTVLVCLLSLPNISSSKAINNNAKYYGLYISGOYKPSVSFNSFKVET	60
Db	1	MNSKRFPTICTSLICLLSPNTSLNFIQNST-KHSGLYVSGHYKPSVIFNSFKVET	59
Qy	61	NVITKNIALKKQVDSIETK-TDASVGISNPSNFTPIPTAVFQDQNSVNFNGTIGYTPAEG	119
Db	60	NHTVQLVALKQVDSIETK-TDASVGISNPSNFTPIPTAVFQDQNSVNFNGTIGYTPAEG	119
Qy	120	TRVEIEGSEYEEFDKQNGPGGYTLSDAYRYPALAREM-----KGNSTFTPEKVS-NSIFHTV	173
Db	120	LNIEVEGSEYEEFDKQNGPGGYTLSDAYRYPALAREM-----KGNSTFTPEKVS-NSIFHTV	179
Qy	174	MRNDGLSIISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDVLHIKPAYQSKLGIAYS	233
Db	180	MRNGLSILSIMINGCYNPLNDLSISPYICGGAGVDAIEFFDVLHIKPAYQSKLGIAYS	239
Qy	234	LPSNISLFAVYHVKMGQNFKNLVQHVLAELASIPKITSAVATLNIGYFGGEIGARLT	292
Db	240	LSDNISLFTNGYHQVIGDQFNKLVQYIGELKENPKITSAVATLNIGYFGGEIGARLT	298

## RESULT 25

US-10-059-964-46  
; Sequence 46, Application US/10059964  
; Patent No. US20020120115A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20020120115A1io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/059,964  
; CURRENT FILING DATE: 2002-01-28  
; EARLIER APPLICATION NUMBER: 09/314,701  
; EARLIER FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 46  
; LENGTH: 299  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
; US-10-059-964-46

Query Match 61.4%; Score 918; DB 12; Length 299;  
Best Local Similarity 60.5%; Pred. No. 4.7e-75;  
Matches 181; Conservative 41; Mismatches 69; Indels 8; Gaps 4;

Qy	1	MNNKLFPTIINTVLVCLLSLPNISSSKAINNNAKYYGLYISGOYKPSVSFNSFKVET	60
Db	1	MNSKRFPTICTSLICLLSPNTSLNFIQNST-KHSGLYVSGHYKPSVIFNSFKVET	59
Qy	61	NVITKNIALKKQVDSIETK-TDASVGISNPSNFTPIPTAVFQDQNSVNFNGTIGYTPAEG	119
Db	60	NHTVQLVALKQVDSIETK-TDASVGISNPSNFTPIPTAVFQDQNSVNFNGTIGYTPAEG	119
Qy	120	TRVEIEGSEYEEFDKQNGPGGYTLSDAYRYPALAREM-----KGNSTFTPEKVS-NSIFHTV	173
Db	120	LNIEVEGSEYEEFDKQNGPGGYTLSDAYRYPALAREM-----KGNSTFTPEKVS-NSIFHTV	179
Qy	174	MRNDGLSIISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDVLHIKPAYQSKLGIAYS	233
Db	180	MRNGLSILSIMINGCYNPLNDLSISPYICGGAGVDAIEFFDVLHIKPAYQSKLGIAYS	239
Qy	234	LPSNISLFAVYHVKMGQNFKNLVQHVLAELASIPKITSAVATLNIGYFGGEIGARLT	292
Db	240	LSDNISLFTNGYHQVIGDQFNKLVQYIGELKENPKITSAVATLNIGYFGGEIGARLT	298

## RESULT 26

US-10-314-639-60  
; Sequence 60, Application US/10314639  
; Publication No. US2003010391A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US2003010391A1  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; TITLE OF INVENTION: Chaffensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/314,639  
; CURRENT FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US/09/314,701  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 60  
; LENGTH: 279  
; TYPE: PRT  
; ORGANISM: P30-11  
US-10-314-639-60

Query Match 52.8%; Score 789.5; DB 9; Length 279;  
Best Local Similarity 55.0%; Pred. No. 1.8e-63;  
Matches 160; Conservative 34; Mismatches 84; Indels 13; Gaps 4;  
Qy 3 NKLKFTIINTVLVCLLSLPNIISSSKAINNAKYYGLYISQYKPSVSVFNSFKETNV 62  
Db 2 NKKKIITVGTTLAYLLLSPLNIISSSKAINNAKYYGLYISQYKPSVSVFNSFKETNV 61  
Qy 63 ITKNLIALKKVDSTETKTDSVGSINPSTFTPIPTAVFQDQNSVNFNGTIGYTFPAEGTRV 122  
Db 62 FTKALIGLRH--DAISTK-----NLTTWDFNTLYKVTQNNIISFSGAIGVSDSTGVRF 114  
Qy 123 EIEGSYEEDVKNPGGYTLSDAYRYFALAREMKNSTPPEKVSNSIFHTVMRNDGLSII 182  
Db 115 EIEGSYEEDVDPGDCIIKDITRYFALARKTSNGH--PNDNGE---YTVMRNDGVSI 168  
Qy 183 SVIVNVCYDFSLNLSISPIYCGGAGVDAIEFFDVLHIKPAYQSKLGIAYSLPNSILFA 242  
Db 169 SVIFNGCYDLSLKELEISPIYCVIGIGDPIEFDDALHIKPAYQSKLGIAYSLPNSILFA 228  
Qy 243 SLVYHKVMGNQFNKLVQHVLAELASIPKITSAVATLNIGYFGGIGARLTF 293  
Db 229 DCYHVRVIGNQFNKLVQHVLAELASIPKITSAVATLNIGYFGGIGARLTF 279

RESULT 27  
US-10-059-964-60  
; Sequence 60, Application US/10059964  
; Patent No. US20020120115A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20020120115A1  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; TITLE OF INVENTION: Chaffensis  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/059,964  
; CURRENT FILING DATE: 2002-01-28  
; EARLIER APPLICATION NUMBER: 09/314,701  
; EARLIER FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 60  
; LENGTH: 279  
; TYPE: PRT  
; ORGANISM: P30-11  
US-10-059-964-60

Query Match 52.8%; Score 789.5; DB 12; Length 279;  
Best Local Similarity 55.0%; Pred. No. 1.8e-63;  
Matches 160; Conservative 34; Mismatches 84; Indels 13; Gaps 4;  
Qy 3 NKLKFTIINTVLVCLLSLPNIISSSKAINNAKYYGLYISQYKPSVSVFNSFKETNV 62

Db 2 NKKKIITVGTTLAYLLLSPLNIISSSKAINNAKYYGLYISQYKPSVSVFNSFKETNV 61  
Qy 63 ITKNLIALKKVDSTETKTDSVGSINPSTFTPIPTAVFQDQNSVNFNGTIGYTFPAEGTRV 122  
Db 62 FTKALIGLRH--DAISTK-----NLTTWDFNTLYKVTQNNIISFSGAIGVSDSTGVRF 114  
Qy 123 EIEGSYEEDVKNPGGYTLSDAYRYFALAREMKNSTPPEKVSNSIFHTVMRNDGLSII 182  
Db 115 EIEGSYEEDVDPGDCIIKDITRYFALARKTSNGH--PNDNGE---YTVMRNDGVSI 168  
Qy 183 SVIVNVCYDFSLNLSISPIYCGGAGVDAIEFFDVLHIKPAYQSKLGIAYSLPNSILFA 242  
Db 169 SVIFNGCYDLSLKELEISPIYCVIGIGDPIEFDDALHIKPAYQSKLGIAYSLPNSILFA 228  
Qy 243 SLVYHKVMGNQFNKLVQHVLAELASIPKITSAVATLNIGYFGGIGARLTF 293  
Db 229 DCYHVRVIGNQFNKLVQHVLAELASIPKITSAVATLNIGYFGGIGARLTF 279

RESULT 28  
US-10-284-986-9  
; Sequence 9, Application US/10284986  
; Publication No. US20030091588A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: Yu, Xu-Jie  
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane  
; TITLE OF INVENTION: Protein Multigene Family  
; FILE REFERENCE: D6311D1  
; CURRENT APPLICATION NUMBER: US/10/284,986  
; CURRENT FILING DATE: 2002-10-31  
; PRIOR APPLICATION NUMBER: 09/846,808  
; PRIOR FILING DATE: 2001-05-01  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 9  
; LENGTH: 285  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: P28-9 Outer Membrane Protein of  
; OTHER INFORMATION: Ehrlichia chaffeensis  
US-10-284-986-9

Query Match 52.7%; Score 789; DB 9; Length 285;  
Best Local Similarity 54.1%; Pred. No. 2e-63;  
Matches 159; Conservative 48; Mismatches 77; Indels 10; Gaps 4;  
Qy 1 MNKLKFTIIN-TVLVCLLSLPNIISSSKAINNAKYYGLYISQYKPSVSVFNSFKVKE 59  
Db 1 MNKRKSFIIIGASLLASLLFTSEASTGNVSNHTYFKPRLYISQYKPSVSHFSKFSVKE 60  
Qy 60 TNVITKNLIALKKVDSTETKTDSVGSINPSTFTPIPTAVFQDQNSVNFNGTIGYTFPAEG 119  
Db 61 TNVNTQLVGLKKDISVI-----GNSNITVTNENFPVIAEFQDNALISFSGAIGVLYSEN 115  
Qy 120 TRVEIEGSYEEDVKNPGGYTLSDAYRYFALAREMKNSTPPEKVSNSIFHTVMRNDGL 179  
Db 116 FRIEVEASYEEDVKNPEG-SATDAYRYFALARAMDG---TNKSSPDTRKFTVMRNDGL 171  
Qy 180 SIISVIVNVCYDFSLNLSISPIYCGGAGVDAIEFFDVLHIKPAYQSKLGIAYSLPNSIS 239  
Db 172 SISSVMINGCYNFTLDDIPVVPYVCAGIGGDFIEFFDNLHVRFAHQKVGIGVSIPEVS 231  
Qy 240 LFASLYHKVMGNQFNKLVQHVLAELASIPKITSAVATLNIGYFGGIGARLTF 293  
Db 232 LFLNGYHKVTGNRFKNLVQHVQVSDLSAPKTSATVATLNIGYFGGIGVRFIF 285

RESULT 29  
US-09-846-808-9  
; Sequence 9, Application US/09846808  
; Patent No. US20020064531A1



```

; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane
; TITLE OF INVENTION: Protein Multigene Family
; FILE REFERENCE: D6311
; CURRENT APPLICATION NUMBER: US/09/846,808
; CURRENT FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 60/201,035
; PRIOR FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 9
; LENGTH: 285
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: P28-9 Outer Membrane Protein of
; OTHER INFORMATION: Ehrlichia chaffeensis
US-09-846-808-9

Query Match          52.7%; Score 789; DB 10; Length 285;
Best Local Similarity 54.1%; Pred. No. 2e-63;
Matches 159; Conservative 48; Mismatches 77; Indels 10; Gaps 4;

Qy 1 MNWKLFTIIN-TVLVCLLSLPNLTSSSKAINNNAKKYGLYISQYKPSVSNFSVKE 59
Db 1 MNRKSFIIIGASLLASLTTSASSTGNVSHTYFKPRLYISQYRPGVSHFSKFSVKE 60
Qy 60 TNVITKNLIALKDVDSIETKTDAISVGSINPSNFTIPYAVFQDINSVNFNGTIGYTPAEG 119
Db 61 TNYNTQVLGUKDLSV-----GNSNITVTNFPYIAEFQDNLASFSGAIGLYSEN 115
Qy 120 TRVIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIPHTVMRNDGL 179
Db 116 FRIEVEASYEEDVKNPEG-SATDAYRYFALARWDG---TNKSPDDTRKFTVMRNDGL 171
Qy 180 SIISVIVNVCYDFSLNLSISPYICGGAGVDAIEFFDLVHLKFAVQSKGLAYSLSNIS 239
Db 172 SISSVMINGCYNFTLDDIPVVPYVVCAGTGGDFIEFFDLVHLKFAHQKVGISISPEVS 231
Qy 240 LFASLYYHKVMGNQFKNLNVQHVLAELASIPKITSAVATLNGYFGGEIGARLTF 293
Db 232 LFLNGYHKVTGNRPFKNLHVQVSDLSAPKFTSAVATLNGYFGGEIGRFTF 285

RESULT 30
US-10-314-639-24
; Sequence 24, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/314,701
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 24
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-10-059-964-24

Query Match          52.6%; Score 786.5; DB 12; Length 279;
Best Local Similarity 54.6%; Pred. No. 3.3e-63;
Matches 160; Conservative 39; Mismatches 77; Indels 17; Gaps 4;

Qy 3 NKLKFTIINTVLVCLLSLPNLTSSSKAINNN--AKKYGLYISQYKPSVSNFSVKE 60
Db 2 SKKFTITIGTVLASLTSFLSIESFSAINHNHTGNTSGIYITGQYRPGVSHFSVKE 61
Qy 61 NVITKNLIALKDVDSIETKTDAISVGSINPSNFTIPYAVFQDINSVNFNGTIGYTPAEG 120
Db 62 NVDTIQLVGKKSASSIDPNT-----YSNFQGPYTVTFQDNLASFSGAIGYSPEL 113
Qy 121 RVIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIPHTVMRNDGLS 180
Db 114 RLELEGSYEKFDVKDPKDYSAKDAFRFFALAR-----NTST---TVPDAQYKTYVMKNGLS 166
Qy 181 IISVIVNVCYDFSLNLSISPYICGGAGVDAIEFFDLVHLKFAVQSKGLAYSLSNIS 240
Db 167 VASIMINGCYDLSFNVLVSPYICAGIGEDFIEFFDLVHLKFAVQSKGLAYSLSNIS 226
Qy 241 FASLYYHKVMGNQFKNLNVQHVLAELASIPKITSAVATLNGYFGGEIGARLTF 293
Db 227 FAGGYHRVIGNPKNLNVNHHVTLDEFPKATSAVATLNVAYFGGEAGVKFTF 279

RESULT 31
US-10-059-964-24
; Sequence 24, Application US/10059964
; Patent No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; CURRENT FILING DATE: 2002-01-28
; EARLIER APPLICATION NUMBER: 09/314,701
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 24
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-10-059-964-24

Query Match          52.6%; Score 786.5; DB 12; Length 279;
Best Local Similarity 54.6%; Pred. No. 3.3e-63;
Matches 160; Conservative 39; Mismatches 77; Indels 17; Gaps 4;

Qy 3 NKLKFTIINTVLVCLLSLPNLTSSSKAINNN--AKKYGLYISQYKPSVSNFSVKE 60
Db 2 SKKFTITIGTVLASLTSFLSIESFSAINHNHTGNTSGIYITGQYRPGVSHFSVKE 61
Qy 61 NVITKNLIALKDVDSIETKTDAISVGSINPSNFTIPYAVFQDINSVNFNGTIGYTPAEG 120
Db 62 NVDTIQLVGKKSASSIDPNT-----YSNFQGPYTVTFQDNLASFSGAIGYSPEL 113
Qy 121 RVIEGSYEEDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIPHTVMRNDGLS 180
Db 114 RLELEGSYEKFDVKDPKDYSAKDAFRFFALAR-----NTST---TVPDAQYKTYVMKNGLS 166
Qy 181 IISVIVNVCYDFSLNLSISPYICGGAGVDAIEFFDLVHLKFAVQSKGLAYSLSNIS 240
Db 167 VASIMINGCYDLSFNVLVSPYICAGIGEDFIEFFDLVHLKFAVQSKGLAYSLSNIS 226
Qy 241 FASLYYHKVMGNQFKNLNVQHVLAELASIPKITSAVATLNGYFGGEIGARLTF 293
Db 227 FAGGYHRVIGNPKNLNVNHHVTLDEFPKATSAVATLNVAYFGGEAGVKFTF 279

RESULT 32
US-10-314-639-30
; Sequence 30, Application US/10314639
; Publication No. US20030103991A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. US20030103991A110

```

;; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
;; FILE REFERENCE: Chaffeensis  
;; CURRENT APPLICATION NUMBER: US/10/314,639  
;; CURRENT FILING DATE: 2002-12-09  
;; PRIOR APPLICATION NUMBER: US/09/314,701  
;; PRIOR FILING DATE: 1999-05-19  
;; NUMBER OF SEQ ID NOS: 66  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 30  
;; LENGTH: 285  
;; TYPE: PRT  
;; ORGANISM: Ehrlichia chaffeensis  
US-10-314-639-30

Query Match 52.4%; Score 784; DB 9; Length 285;  
Best Local Similarity 53.7%; Pred. No. 5.7e-63;  
Matches 158; Conservative 48; Mismatches 78; Indels 10; Gaps 4;  
QY 1 MNKLKFTIIN-TVLVCLLSLSPNISSKAINNAKKYGLYISGOYKPSVSNFSVKE 59  
DB 1 MNKRKSFLLIGASLASLFTSEASTGNVSNHTYFKPRLYISGOYRPGVSHFSVKE 60  
QY 60 TNVITKNLIALKQVDSIETKTDAISVGNPSNFTIPYTAVDQNSVNFNGTIGYTFAG 119  
DB 61 TNYNTQLVGLKDISI-----GNSNITTYTNFNPPIAEFQDVAISFGAIGLYSEN 115  
QY 120 TRVEIGSYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGL 179  
DB 116 FRIEVEASYEEFDVKNPEG-SATDAYRYFALARAMDG---TNKSSPDTRKFTVMRNDGL 171  
QY 180 SIISVIVNVCYDFSLNNLSISPIYICGGAGVDAIEFDVLHIKFAYSQKLGIAYSLPNIS 239  
DB 172 SISSWMINGCYNFTLDDIPVVPYVACAGIGGDFIEFFNDLHVKFRHQKVGISYSISPEVS 231  
QY 240 LFASLYYHKVMGNQFKNLVQHVAELASIPKITSAVATLNIGYFGGEGICARLTF 293  
DB 232 LFLNGYHYKVTGNRFKNLVQHVSVDLSADPKFTSAVATLNIGYFGGEGICARLTF 285  
QY 241 FASLYYHKVMGNQFKNLVQHVAELASIPKITSAVATLNIGYFGGEGICARLTF 293  
DB 227 FAGGYHVRVIGNKFNLVNHNVTTPDEFKATSAVATLNIGYFGGEGICARLTF 279

RESULT 33  
US-10-059-964-30  
;; Sequence 30, Application US/10059964  
;; Patent No. US20020120115A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Rikihisa, Yasuko  
;; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
;; FILE REFERENCE: 22727/04021  
;; CURRENT APPLICATION NUMBER: US/10/059,964  
;; CURRENT FILING DATE: 2002-01-28  
;; EARLIER APPLICATION NUMBER: 09/314,701  
;; EARLIER FILING DATE: 1999-05-19  
;; NUMBER OF SEQ ID NOS: 66  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 30  
;; LENGTH: 285  
;; TYPE: PRT  
;; ORGANISM: Ehrlichia chaffeensis  
US-10-059-964-30

Query Match 52.4%; Score 784; DB 12; Length 285;  
Best Local Similarity 53.7%; Pred. No. 5.7e-63;  
Matches 158; Conservative 48; Mismatches 78; Indels 10; Gaps 4;  
QY 1 MNKLKFTIIN-TVLVCLLSLSPNISSKAINNAKKYGLYISGOYKPSVSNFSVKE 59  
DB 1 MNKRKSFLLIGASLASLFTSEASTGNVSNHTYFKPRLYISGOYRPGVSHFSVKE 60  
QY 60 TNVITKNLIALKQVDSIETKTDAISVGNPSNFTIPYTAVDQNSVNFNGTIGYTFAG 119  
DB 61 TNYNTQLVGLKDISI-----GNSNITTYTNFNPPIAEFQDVAISFGAIGLYSEN 115

QY 120 TRVEIGSYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGL 179  
DB 116 FRIEVEASYEEFDVKNPEG-SATDAYRYFALARAMDG---TNKSSPDTRKFTVMRNDGL 171  
QY 180 SIISVIVNVCYDFSLNNLSISPIYICGGAGVDAIEFDVLHIKFAYSQKLGIAYSLPNIS 239  
DB 172 SISSWMINGCYNFTLDDIPVVPYVACAGIGGDFIEFFNDLHVKFRHQKVGISYSISPEVS 231  
QY 240 LFASLYYHKVMGNQFKNLVQHVAELASIPKITSAVATLNIGYFGGEGICARLTF 293  
DB 232 LFLNGYHYKVTGNRFKNLVQHVSVDLSADPKFTSAVATLNIGYFGGEGICARLTF 285

RESULT 34  
US-10-284-986-6  
;; Sequence 6, Application US/10284986  
;; Publication No. US20030091588A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Walker, David H.  
;; APPLICANT: Yu, Xu-Jie  
;; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane  
;; FILE REFERENCE: D6311D1  
;; CURRENT APPLICATION NUMBER: US/10/284,986  
;; CURRENT FILING DATE: 2002-10-31  
;; PRIOR APPLICATION NUMBER: 09/846,808  
;; PRIOR FILING DATE: 2001-05-01  
;; NUMBER OF SEQ ID NOS: 53  
;; SEQ ID NO 6  
;; LENGTH: 279  
;; TYPE: PRT  
;; ORGANISM: Ehrlichia chaffeensis  
;; FEATURE:  
;; OTHER INFORMATION: P28-6 Outer Membrane Protein of  
;; OTHER INFORMATION: Ehrlichia chaffeensis  
US-10-284-986-6

Query Match 52.1%; Score 779.5; DB 9; Length 279;  
Best Local Similarity 54.3%; Pred. No. 1.4e-62;  
Matches 159; Conservative 39; Mismatches 78; Indels 17; Gaps 4;  
QY 3 NKLKFTIINTVLVCLLSLSPNISSKAINNN--AKKYGLYISGOYKPSVSNFSVKE 60  
DB 2 SKKFTITIGTVLASLFLSIESFSAINHNHTGNTSGIYITGYRPGVSHFSVKE 61  
QY 61 NVITKNLIALKQVDSIETKTDAISVGNPSNFTIPYTAVDQNSVNFNGTIGYTFAG 120  
DB 62 NVDTIQLVGYKKSASSIDPNT-----YSNFGPPTVTTFQDVAASFGAIGYSYPSL 113  
QY 121 RVEIGSYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGLS 180  
DB 114 RLEEGSYEKFDVKNPKDYSAKDAFRFFALAR----NTST---TVPDAQYTVNKNGLS 166  
QY 181 IISVIVNVCYDFSLNNLSISPIYICGGAGVDAIEFDVLHIKFAYSQKLGIAYSLPNISL 240  
DB 167 VASIMINGCYDLSFPNNLVSPYICAGIGDFIEFFDTHLKLAYQKLGISYFFPKIN 226  
QY 241 FASLYYHKVMGNQFKNLVQHVAELASIPKITSAVATLNIGYFGGEGICARLTF 293  
DB 227 FAGGYHVRVIGNKFNLVNHNVTTPDEFKATSAVATLNIGYFGGEGICARLTF 279

RESULT 35  
US-09-846-808-6  
;; Sequence 6, Application US/09846808  
;; Patent No. US20020064531A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Walker, David H.  
;; APPLICANT: Yu, Xu-Jie  
;; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane  
;; FILE REFERENCE: D6311

```
; CURRENT APPLICATION NUMBER: US/09/846,808
; CURRENT FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 60/201,035
; PRIOR FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 6
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: P28-6 Outer Membrane Protein of
; OTHER INFORMATION: Ehrlichia chaffeensis
US-09-846-808-6

Query Match      52.1%; Score 779.5; DB 10; Length 279;
Best Local Similarity 54.3%; Pred. No. 1.4e-62;
Matches 159; Conservative 39; Mismatches 78; Indels 17; Gaps 4;

Qy  3 NKLKFTIINTVLCLLSLPNIISSSKAINNN--AKYGYGLYISGQYKPSVSVFNSFV 60
Db  2 SKKPFITIGTVLALLSLPISLSEFSAINHHNHTGNTSGIYITGQYRPGVSHFSNFVKET 61
Qy  61 NVITKNLIALKKVDISITKTDAVSGISNPSNFTIPYTAVFQDNSVNFNGTIGYTFAGT 120
Db  62 NVDTIQLVGYKKSASIDPNT-----YNFQGPYTVTFQDAAESFGAIGSYPESL 113
Qy  121 RVEIEGSYEEDVKNPGGYTLSDAYRYPALAREMKGNSFTPKPKVSNISFIHTVMRNDGLS 180
Db  114 RLELEGSYEEDVKNPKDYSADAFRRPALAR----NTST--TVPDQAQYVMKNGLS 166
Qy  181 IISVIVNVCYDFSLNLSISPYICGGAGVDAIEFDFVLHIKFAVQSKLGIAYSLPSN 240
Db  167 VASIMINGCYDLSFNNLVSPYICAGIGEDFIEFDFTLHIKLAQYQKLGISYFPFKIN 226
Qy  241 FASLYYHKMGQNPKNLVQHVLAELASPKITSATATINIGYFGEIGARLTF 293
Db  227 FAGGYHVRVIGNKFNLVNHTVTFDEPKATSAVATLVNAYFGEAGVKTF 279

RESULT 36
US-10-284-986-7
; Sequence 7, Application US/10284986
; Publication No. US20030091588A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane
; TITLE OF INVENTION: Protein Multigene Family
; FILE REFERENCE: D6311D1
; CURRENT APPLICATION NUMBER: US/10/284,986
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: 09/846,808
; PRIOR FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 7
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: P28-7 Outer Membrane Protein of
; OTHER INFORMATION: Ehrlichia chaffeensis
US-10-284-986-7

Query Match      51.3%; Score 768; DB 9; Length 283;
Best Local Similarity 54.4%; Pred. No. 1.6e-61;
Matches 161; Conservative 41; Mismatches 78; Indels 16; Gaps 6;

Qy  1 MNKCLKFTIINTVLCLLS-LP--NISSKAINNNAKYGYGLYISGQYKPSVSVFNSFV 57
Db  1 MSACKKLFIIIGSVLVCLVSLPTKSLNLNNINNT--KCTGLYVSGQYKPTVSHFSNPSL 59
Qy  58 KETNVITKNLIALKKVDISITKTDAVSGISNPSNFTIPYTAVFQDNSVNFNGTIGYTF 117
Db  60 KETVDTKELLGLAKDIKSI---TD---ITTNNKFNIPYNTKFDQNAVSAAGVYISQ 112
Qy  118 EGTREIEGSYEEDVKNPGGYTLSDAYRYPALAREMKGNSFTPKPKVSNISFIHTVMRND 177
Db  113 DSPRVEVSWSEYEDVKNPGYVYVSEAFRYALARGIDN-----LQKYPETNKYVVIKN 167
Qy  178 GLSIISVIVNVCYDFSLNLSISPYICGGAGVDAIEFDFVLHIKFAVQSKLGIAYSLPSN 237
Db  168 GLSVASIIINGCYDLSLNNLKVSPIYICVFGGDIIEFFSAVSPKFAVQYGVISYPLFSN 227
Qy  238 ISLFSALYYHKMGQNPKNLVQHVLAELASPKITSATATINIGYFGEIGARLTF 293
Db  228 MIIPADGYHKVIGNKFNLVNHTVTFDEPKATSAVATLVNAYFGEAGVKTF 283

RESULT 38
US-09-846-808-7
; Sequence 7, Application US/09846808
; Patent No. US20020064531A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xu-Jie
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane
; TITLE OF INVENTION: Protein Multigene Family
; FILE REFERENCE: D6311
; CURRENT APPLICATION NUMBER: US/09/846,808

Qy  60 KETVDTKELLGLAKDIKSI---TD---ITTNNKFNIPYNTKFDQNAVSAAGVYISQ 112
Qy  118 EGTREIEGSYEEDVKNPGGYTLSDAYRYPALAREMKGNSFTPKPKVSNISFIHTVMRND 177
Db  113 DSPRVEVSWSEYEDVKNPGYVYVSEAFRYALARGIDN-----LQKYPETNKYVVIKN 167
Qy  178 GLSIISVIVNVCYDFSLNLSISPYICGGAGVDAIEFDFVLHIKFAVQSKLGIAYSLPSN 237
Db  168 GLSVASIIINGCYDLSLNNLKVSPIYICVFGGDIIEFFSAVSPKFAVQYGVISYPLFSN 227
Qy  238 ISLFSALYYHKMGQNPKNLVQHVLAELASPKITSATATINIGYFGEIGARLTF 293
Db  228 MIIPADGYHKVIGNKFNLVNHTVTFDEPKATSAVATLVNAYFGEAGVKTF 283
```

;; CURRENT FILING DATE: 2001-05-01  
;; PRIOR APPLICATION NUMBER: 60/201,035  
;; PRIOR FILING DATE: 2000-05-01  
;; NUMBER OF SEQ ID NOS: 53  
;; SEQ ID NO 7  
;; LENGTH: 283  
;; TYPE: PRT  
;; ORGANISM: Ehrlichia chaffeensis  
;; FEATURE:  
;; OTHER INFORMATION: P28-7 Outer Membrane Protein of  
;; OTHER INFORMATION: Ehrlichia chaffeensis  
US-09-846-808-7

Query Match 51.3%; Score 768; DB 10; Length 283;  
Best Local Similarity 54.4%; Pred. No. 1.6e-61;  
Matches 161; Conservative 41; Mismatches 78; Indels 16; Gaps 6;  
QY 1 MNKKLFTIINTVLVCLLS-LP--NISSKAINNNKAKYGLYISQYKPSVSVFNSFV 57  
Db 1 MSAKKLFIIIGSVLVCLVSLYPTKLSLNINNNNT-KCTGLYVSGQYKPTVSHFNSFL 59  
QY 58 KETNVITKNLIALKQVDSEIETKTDASVGSINPSNFTIPYTAVFQDNSVNFNGTIGYTFA 117  
Db 60 KETVTDTRKLLGLAKDIKSI---TD---ITNKKFNIPYNTKFDQNAVSAVSAAGVYSIQ 112  
QY 118 EGTREIEGSEYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRND 177  
Db 113 DSPRVEEWSYEEFDVKNPGVNVVSEAFRYIALARGIDN-----LQKYPETNKYVVIKN 167  
QY 178 GLSIISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDVLHIKPAYOSKLGIAYSPLSN 237  
Db 168 GLSVASIIINGCYDFSLNNLKVSPIYICVGGGDIIEFFSAVSPKFAQKVGISYPLFSN 227  
QY 238 ISLFASLYYHKMGNOFKNLVQHVLAELASIPKITSAVATLNIGYFGGEIGARLTF 293  
Db 228 MIIPADGYHYHKVGNKFNLLNVQHVSLNSHPKSTFAVATLNVEYFGSEFGLKPIF 283

RESULT 39  
US-10-059-964-26  
; Sequence 26, Application US/10059964  
; Patent No. US20020120115A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20020120115a1io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/059,964  
; CURRENT FILING DATE: 2002-01-28  
; EARLIER APPLICATION NUMBER: 09/314,701  
; EARLIER FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 26  
; LENGTH: 283  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
US-10-059-964-26

Query Match 51.3%; Score 768; DB 12; Length 283;  
Best Local Similarity 54.4%; Pred. No. 1.6e-61;  
Matches 161; Conservative 41; Mismatches 78; Indels 16; Gaps 6;  
QY 1 MNKKLFTIINTVLVCLLS-LP--NISSKAINNNKAKYGLYISQYKPSVSVFNSFV 57  
Db 1 MSAKKLFIIIGSVLVCLVSLYPTKLSLNINNNNT-KCTGLYVSGQYKPTVSHFNSFL 59  
QY 58 KETNVITKNLIALKQVDSEIETKTDASVGSINPSNFTIPYTAVFQDNSVNFNGTIGYTFA 117  
Db 60 KETVTDTRKLLGLAKDIKSI---TD---ITNKKFNIPYNTKFDQNAVSAVSAAGVYSIQ 112  
QY 118 EGTREIEGSEYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRND 177

Db 113 DSPRVEEWSYEEFDVKNPGVNVVSEAFRYIALARGIDN-----LQKYPETNKYVVIKN 167  
QY 178 GLSIISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDVLHIKPAYOSKLGIAYSPLSN 237  
Db 168 GLSVASIIINGCYDFSLNNLKVSPIYICVGGGDIIEFFSAVSPKFAQKVGISYPLFSN 227  
QY 238 ISLFASLYYHKMGNOFKNLVQHVLAELASIPKITSAVATLNIGYFGGEIGARLTF 293  
Db 228 MIIPADGYHYHKVGNKFNLLNVQHVSLNSHPKSTFAVATLNVEYFGSEFGLKPIF 283

RESULT 40  
US-10-314-639-58  
; Sequence 58, Application US/10314639  
; Publication No. US20030103991A1  
; GENERAL INFORMATION:  
; APPLICANT: Rikihisa, Yasuko  
; APPLICANT: Ohasi, No. US20030103991A1io  
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia  
; FILE REFERENCE: 22727/04021  
; CURRENT APPLICATION NUMBER: US/10/314,639  
; CURRENT FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US/09/314,701  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 58  
; LENGTH: 281  
; TYPE: PRT  
; ORGANISM: p30-9  
US-10-314-639-58

Query Match 49.5%; Score 740; DB 9; Length 281;  
Best Local Similarity 52.2%; Pred. No. 5.3e-59;  
Matches 153; Conservative 41; Mismatches 87; Indels 12; Gaps 3;  
QY 1 MNKKLFTIINTVLVCLLSLPNIISSKAINNNKAKYGLYISQYKPSVSVFNSFVKET 60  
Db 1 MNKKRNFPLIGMSLLINLLLPIDASSMEVHNHYTHFTPRLYISQYRPGVSHFSKFSVKET 60  
QY 61 NVITKNLIALKQVDSEIETKTDASVGSINPSNFTIPYTAVFQDNSVNFNGTIGYTFAEGT 120  
Db 61 HCNVQLVGLTKDI-----KVTNNSINTNTSFNFPVFAEFQDMSFSGAIGCFYSHF 115  
QY 121 RVEIEGSEYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPKKVSNSIFHTVMRNDGLS 180  
Db 116 RIEVEASVEEFDVKNPEGST--TDSYRYFALARGMDGNIPTSQK-----FTVMRNDGLL 168  
QY 181 IISVIVNVCYDFSLNNLSISPYICGGAGVDAIEFFDVLHIKPAYOSKLGIAYSPLSNISL 240  
Db 169 ISSVMINGCVNVLNDIAQAEPIYICAGLGGDFIEFNGFHVKLAYQKVGISYQIFPEVRL 228  
QY 241 FASLYYHKMGNOFKNLVQHVLAELASIPKITSAVATLNIGYFGGEIGARLTF 293  
Db 229 FIDGYHYHKVGNKFNLLNVQHVLAELASIPKITSAVATLNIGYFGGEIGARLTF 281

Search completed: July 4, 2003, 06:41:22  
Job time : 241 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 4, 2003, 05:14:45 ; Search time 156 Seconds  
(without alignments)  
386.999 Million cell updates/sec

Title: US-10-062-624-40

Perfect score: 293

Sequence: 1 MNKKLKEITIVLVCLLSL.....AVATNLGYGGEIGARLTF 293

Scoring table:

OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

SPTREMBL 21:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_muc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	293	100.0	293	2	Q9f477 ehrlichia c
2	25	8.5	288	2	Q9L6V1 ehrlichia c
3	25	8.5	297	2	Q52102 ehrlichia c
4	24	8.2	296	2	Q9ADV5 ehrlichia c
5	22	7.5	298	2	Q9L6V3 ehrlichia c
6	17	5.8	300	2	Q9L6V2 ehrlichia c
7	16	5.5	291	2	Q9L6V4 ehrlichia c
8	16	5.5	291	2	Q9ACJ0 ehrlichia c
9	15	5.1	275	2	Q9L6V6 ehrlichia c
10	15	5.1	289	2	Q9ADV9 ehrlichia c
11	13	4.4	281	2	Q9ADV7 ehrlichia c
12	11	3.8	279	2	Q9L6V9 ehrlichia c
13	11	3.8	279	2	Q9ACJ2 ehrlichia c
14	11	3.8	283	2	Q9L6V7 ehrlichia c
15	11	3.8	295	2	Q9L6Z0 ehrlichia c
16	11	3.8	299	2	Q9ADV6 ehrlichia c

17	10	3.4	285	2	Q9L6V5
18	10	3.4	294	2	Q9ADV4
19	9	3.1	278	2	Q9ADM1
20	9	3.1	279	2	Q9ADV8
21	8	2.7	98	17	Q8UOV7
22	8	2.7	265	2	Q9AF99
23	8	2.7	268	2	Q9B63
24	8	2.7	268	2	Q9B61
25	8	2.7	269	2	Q9B62
26	8	2.7	270	2	Q9AF98
27	8	2.7	271	2	Q9AFA0
28	8	2.7	272	2	Q9AMF6
29	8	2.7	272	2	Q9B54
30	8	2.7	272	2	Q9L6Z1
31	8	2.7	275	2	Q9B59
32	8	2.7	276	2	Q9B60
33	8	2.7	276	2	Q9B53
34	8	2.7	277	2	Q9B65
35	8	2.7	278	2	Q9B57
36	8	2.7	278	2	Q9B56
37	8	2.7	278	2	Q9B52
38	8	2.7	280	2	Q9F476
39	8	2.7	281	2	Q9S6H0
40	8	2.7	281	2	Q9S6H1
41	8	2.7	281	2	Q46328
42	8	2.7	284	2	Q9AFA1
43	8	2.7	287	2	Q46329
44	8	2.7	287	2	Q9R425
45	8	2.7	290	2	Q46324
46	8	2.7	290	2	Q46330
47	8	2.7	290	2	Q46332
48	8	2.7	290	2	Q9B64
49	8	2.7	290	2	Q46333
50	8	2.7	353	16	Q9WX08
51	8	2.7	495	5	Q8T2K0
52	8	2.7	620	16	Q8YSY0
53	8	2.7	1036	16	Q9L0V7
54	7	2.4	75	2	Q8VTT8
55	7	2.4	93	5	Q21047
56	7	2.4	117	16	Q9K4K9
57	7	2.4	128	5	Q9W305
58	7	2.4	137	8	Q34503
59	7	2.4	142	2	O50597
60	7	2.4	142	2	O50600
61	7	2.4	142	17	Q979X3
62	7	2.4	146	16	Q92XJ4
63	7	2.4	157	16	Q8XJ05
64	7	2.4	187	12	Q8QN58
65	7	2.4	189	8	Q8W7H0
66	7	2.4	197	17	O29094
67	7	2.4	204	17	Q8ZWA1
68	7	2.4	206	13	Q8UWG3
69	7	2.4	223	13	Q8UWK8
70	7	2.4	240	8	Q9TEW2
71	7	2.4	240	8	Q9TEW1
72	7	2.4	244	13	Q8UWL0
73	7	2.4	244	16	Q9HWY9
74	7	2.4	253	10	O49388
75	7	2.4	254	8	Q94X42
76	7	2.4	256	16	Q92LR6
77	7	2.4	256	5	Q8T2J8
78	7	2.4	256	8	Q94X41
79	7	2.4	256	8	Q8WA56
80	7	2.4	257	8	Q953N6
81	7	2.4	257	8	Q953N5
82	7	2.4	257	8	Q953N4
83	7	2.4	257	8	Q8WA74
84	7	2.4	258	8	Q953N2
85	7	2.4	258	8	Q94X39
86	7	2.4	258	8	Q8WA68
87	7	2.4	258	8	Q8WA49
88	7	2.4	259	8	Q95V70
89	7	2.4	259	8	Q94X43

90	7	2.4	259	8	Q94X40	Q94X40 bemisia tab	163	7	2.4	424	5	O61444	O61444 drosophila
91	7	2.4	259	8	Q957V1	Q957V1 bemisia tab	164	7	2.4	433	16	Q928S3	Q928S3 listeria in
92	7	2.4	259	8	Q8WA69	Q8WA69 bemisia tab	165	7	2.4	445	5	Q9GXR9	Q9GXR9 leishmania
93	7	2.4	259	8	Q8WA67	Q8WA67 bemisia tab	166	7	2.4	455	16	Q97FJ2	Q97FJ2 clostridium
94	7	2.4	259	8	Q8WA66	Q8WA66 bemisia tab	167	7	2.4	456	11	P97861	P97861 mus musculus
95	7	2.4	259	8	Q8WA64	Q8WA64 bemisia tab	168	7	2.4	467	10	Q9FF99	Q9FF99 arabisopsis
96	7	2.4	259	8	Q8WA62	Q8WA62 bemisia tab	169	7	2.4	483	4	P78387	P78387 homo sapien
97	7	2.4	259	8	Q8WA61	Q8WA61 bemisia tab	170	7	2.4	486	4	O43790	O43790 homo sapien
98	7	2.4	259	8	Q8WA60	Q8WA60 bemisia tab	171	7	2.4	493	4	P78385	P78385 homo sapien
99	7	2.4	259	8	Q8WA59	Q8WA59 bemisia tab	172	7	2.4	493	4	Q9NSB3	Q9NSB3 homo sapien
100	7	2.4	259	8	Q8WA58	Q8WA58 bemisia tab	173	7	2.4	495	11	O61726	O61726 mus musculus
101	7	2.4	259	8	Q8WA57	Q8WA57 bemisia tab	174	7	2.4	496	10	O8SOM9	O8SOM9 oryza sativ
102	7	2.4	259	8	Q8WA55	Q8WA55 bemisia tab	175	7	2.4	505	4	O14533	O14533 homo sapien
103	7	2.4	259	8	Q8WA54	Q8WA54 bemisia tab	176	7	2.4	506	16	O69862	O69862 streptomyce
104	7	2.4	259	8	Q8WA53	Q8WA53 bemisia tab	177	7	2.4	508	6	Q28582	Q28582 ovis aries
105	7	2.4	259	8	Q8WA50	Q8WA50 bemisia tab	178	7	2.4	511	4	Q9NSB4	Q9NSB4 homo sapien
106	7	2.4	259	8	Q8WA47	Q8WA47 bemisia tab	179	7	2.4	513	17	O9YC36	O9YC36 aeropyrum p
107	7	2.4	259	8	Q8WA46	Q8WA46 bemisia tab	180	7	2.4	517	10	O94BW8	O94BW8 arabisopsis
108	7	2.4	259	8	Q8WA45	Q8WA45 bemisia tab	181	7	2.4	531	17	O97CH8	O97CH8 thermoplasma
109	7	2.4	260	8	Q94X47	Q94X47 bemisia tab	182	7	2.4	533	2	O49521	O49521 mycoplasma
110	7	2.4	260	8	Q94X46	Q94X46 bemisia tab	183	7	2.4	540	13	O98IS7	O98IS7 takifugu ru
111	7	2.4	260	8	Q94X45	Q94X45 bemisia tab	184	7	2.4	551	4	O95678	O95678 homo sapien
112	7	2.4	260	8	Q94X44	Q94X44 bemisia tab	185	7	2.4	551	4	O9NSA9	O9NSA9 homo sapien
113	7	2.4	260	8	Q8WA65	Q8WA65 bemisia tab	186	7	2.4	555	4	O9Y5Y0	O9Y5Y0 homo sapien
114	7	2.4	260	8	Q8WA51	Q8WA51 bemisia tab	187	7	2.4	594	5	O04579	O04579 drosophila
115	7	2.4	261	8	Q8WA63	Q8WA63 bemisia tab	188	7	2.4	600	4	Q9NSB2	Q9NSB2 homo sapien
116	7	2.4	261	8	Q8WA52	Q8WA52 bemisia tab	189	7	2.4	604	11	Q99M73	Q99M73 mus musculus
117	7	2.4	261	16	P95283	P95283 mycobacteri	190	7	2.4	627	2	O93IT3	O93IT3 salmonella
118	7	2.4	268	12	Q89705	Q89705 african swi	191	7	2.4	627	16	O8ZRL9	O8ZRL9 salmonella
119	7	2.4	268	12	O65130	O65130 african swi	192	7	2.4	627	16	O8Z978	O8Z978 salmonella
120	7	2.4	271	17	O97A26	O97A26 thermoplasma	193	7	2.4	635	11	O8QZ22	O8QZ22 mus musculus
121	7	2.4	279	4	Q9NVR9	Q9NVR9 homo sapien	194	7	2.4	637	10	O9SE50	O9SE50 arabisopsis
122	7	2.4	280	2	Q93E58	Q93E58 cowdria rum	195	7	2.4	646	10	O8RVH4	O8RVH4 cichorium i
123	7	2.4	280	2	Q93E55	Q93E55 cowdria rum	196	7	2.4	663	10	O9FJ13	O9FJ13 arabisopsis
124	7	2.4	280	8	Q94UH3	Q94UH3 bemisia tab	197	7	2.4	721	10	O9ZP30	O9ZP30 carica papa
125	7	2.4	280	8	Q94UH1	Q94UH1 bemisia tab	198	7	2.4	762	3	O9UTE0	O9UTE0 schizosacch
126	7	2.4	280	8	Q94Q30	Q94Q30 bemisia tab	199	7	2.4	783	3	O9HE09	O9HE09 schizosacch
127	7	2.4	283	2	Q9F474	Q9F474 ehrlichia c	200	7	2.4	783	16	Q97D91	Q97D91 clostridium
128	7	2.4	284	17	Q96ZQ1	Q96ZQ1 sulfolobus	201	7	2.4	827	10	Q9ZVW8	Q9ZVW8 arabisopsis
129	7	2.4	284	17	Q9HM42	Q9HM42 thermoplasma	202	7	2.4	829	5	P90754	P90754 caenorhabdi
130	7	2.4	289	2	Q9ADM2	Q9ADM2 ehrlichia c	203	7	2.4	851	16	P72594	P72594 synechocyst
131	7	2.4	290	2	Q9AEU3	Q9AEU3 cowdria rum	204	7	2.4	850	12	O88171	O88171 strawberry
132	7	2.4	292	16	Q8ZR83	Q8ZR83 salmonella	205	7	2.4	936	13	O90XA7	O90XA7 brachydanio
133	7	2.4	292	16	Q8Z8Q9	Q8Z8Q9 salmonella	206	7	2.4	974	10	O8S905	O8S905 arabisopsis
134	7	2.4	292	16	Q8XCX6	Q8XCX6 escherichia	207	7	2.4	982	5	Q9VUAS	Q9VUAS drosophila
135	7	2.4	293	16	O8YED2	O8YED2 brucella me	208	7	2.4	1003	10	Q9LFP5	Q9LFP5 arabisopsis
136	7	2.4	303	10	O941B6	O941B6 arabisopsis	209	7	2.4	1013	5	Q9XZ33	Q9XZ33 drosophila
137	7	2.4	310	3	Q03193	Q03193 saccharomyc	210	7	2.4	1024	16	O8ZCV9	O8ZCV9 versinia pe
138	7	2.4	312	16	O8ZQ30	O8ZQ30 salmonella	211	7	2.4	1025	16	O8X7J3	O8X7J3 escherichia
139	7	2.4	312	16	O8Z7M6	O8Z7M6 salmonella	212	7	2.4	1026	16	O8ZNO1	O8ZNO1 salmonella
140	7	2.4	312	16	O8YLD4	O8YLD4 anabaena sp	213	7	2.4	1026	16	O8ZSF6	O8ZSF6 salmonella
141	7	2.4	313	16	O8Y2S3	O8Y2S3 ralbstonia s	214	7	2.4	1050	16	O67202	O67202 aquifex aeo
142	7	2.4	322	16	O97D21	O97D21 clostridium	215	7	2.4	1109	8	O98S72	O98S72 guillardia
143	7	2.4	334	5	O61107	O61107 trypanosoma	216	7	2.4	1213	3	O13818	O13818 schizosacch
144	7	2.4	334	5	O61086	O61086 trypanosoma	217	7	2.4	1340	12	O65177	O65177 african swi
145	7	2.4	334	16	Q9CG08	Q9CG08 lactococcus	218	7	2.4	1349	4	O9P2C7	O9P2C7 homo sapien
146	7	2.4	336	10	Q9XFW8	Q9XFW8 beta vulgar	219	7	2.4	1404	5	Q20073	Q20073 caenorhabdi
147	7	2.4	343	2	O85977	O85977 sphingomona	220	7	2.4	1487	11	O62028	O62028 mus musculus
148	7	2.4	343	2	O8Z3U6	O8Z3U6 sphingomona	221	7	2.4	1639	2	O9X577	O9X577 bacillus su
149	7	2.4	343	5	O9U614	O9U614 leishmania	222	7	2.4	1696	4	O9Y4F4	O9Y4F4 homo sapien
150	7	2.4	343	5	O9VDW5	O9VDW5 drosophila	223	7	2.4	1903	5	O9U5D6	O9U5D6 plautia sta
151	7	2.4	351	5	P91144	P91144 caenorhabdi	224	7	2.4	2090	5	O9M2T1	O9M2T1 drosophila
152	7	2.4	358	17	O29471	O29471 archaeoglob	225	7	2.4	2162	12	O91940	O91940 bovine resp
153	7	2.4	376	5	P90949	P90949 caenorhabdi	226	7	2.4	2162	12	Q9WKU5	Q9WKU5 bovine resp
154	7	2.4	381	11	Q99MH7	Q99MH7 mus musculus	227	7	2.4	2969	4	Q9NR48	Q9NR48 homo sapien
155	7	2.4	384	16	O9HTH4	O9HTH4 pseudomonas	228	7	2.4	3583	2	O45675	O45675 bacillus su
156	7	2.4	388	16	Q9HW46	Q9HW46 pseudomonas	229	6	2.0	20	2	O9R4C0	O9R4C0 actinoplane
157	7	2.4	390	11	Q9ZER2	Q9ZER2 mus musculus	230	6	2.0	28	2	O49327	O49327 mycoplasma
158	7	2.4	395	4	Q96M17	Q96M17 homo sapien	231	6	2.0	32	8	Q36434	Q36434 farfantepe
159	7	2.4	395	4	Q96KC9	Q96KC9 homo sapien	232	6	2.0	38	16	Q8XZJ1	Q8XZJ1 escherichia
160	7	2.4	396	6	Q95JW6	Q95JW6 macaca fasc	233	6	2.0	41	9	O9MC40	O9MC40 bacterioph
161	7	2.4	398	17	Q8ZV87	Q8ZV87 pyrobaculum	234	6	2.0	43	5	O9N6B7	O9N6B7 chironomus
162	7	2.4	410	11	Q8R5A7	Q8R5A7 mus musculus	235	6	2.0	46	12	Q91F83	Q91F83 chilo iride







528	6	2.0	207	12	056649	056649 feline cali	601	233	12	065528	065528 guaroa viru
529	6	2.0	207	16	08XA62	08XA62 escherichia	602	233	16	0983L9	0983L9 rhizobium l
530	6	2.0	208	16	0985J9	0985J9 rhizobium l	603	234	11	035517	035517 rattus sord
531	6	2.0	208	16	097IR3	097IR3 clostridium	604	234	15	09IEC2	09IEC2 human immun
532	6	2.0	209	2	087949	087949 shewanella	605	234	16	09FBW9	09FBW9 streptomyce
533	6	2.0	210	8	08SG42	08SG42 cf. senosto	606	235	5	09N3V8	09N3V8 caenorhabdi
534	6	2.0	210	12	056314	056314 feline cali	607	235	16	092JEC3	092JEC3 rickettsia
535	6	2.0	210	12	056315	056315 feline cali	608	235	16	08YAY5	08YAY5 brucella me
536	6	2.0	210	16	09AC09	09AC09 caulobacter	609	235	17	08TMJ6	08TMJ6 methanosarc
537	6	2.0	211	11	09CRF0	09CRF0 mus musculu	610	236	12	08EMN5	08EMN5 amascta moo
538	6	2.0	211	16	097KY7	097KY7 clostridium	611	236	16	08UBG5	08UBG5 agrobacteri
539	6	2.0	212	2	085372	085372 rhodococcus	612	237	3	09P8B5	09P8B5 agaricus bi
540	6	2.0	212	3	074318	074318 schizosacch	613	237	16	09A1S1	09A1S1 streptococc
541	6	2.0	213	2	09RFE0	09RFE0 rhodobacter	614	240	2	08VVM7	08VVM7 vibrio chol
542	6	2.0	213	5	09GY18	09GY18 leishmania	615	240	4	096LI4	096LI4 homo sapien
543	6	2.0	215	10	09S9C1	09S9C1 nicotiana a	616	240	5	002294	002294 caenorhabdi
544	6	2.0	216	2	09ZFM0	09ZFM0 bacillus st	617	240	5	066727	066727 equine infe
545	6	2.0	216	10	09SUL5	09SUL5 arabidopsis	618	240	15	066728	066728 equine infe
546	6	2.0	216	12	084523	084523 paramedium	619	240	16	09A1B0	09A1B0 streptococc
547	6	2.0	216	16	08YGS4	08YGS4 brucella me	620	240	17	0973B4	0973B4 sulfolobus
548	6	2.0	217	4	096RC4	096RC4 homo sapien	621	241	5	062460	062460 caenorhabdi
549	6	2.0	217	4	096RC4	096RC4 homo sapien	622	241	5	09U0Y0	09U0Y0 leishmania
550	6	2.0	217	5	045970	045970 caenorhabdi	623	241	16	09CL93	09CL93 pasteurella
551	6	2.0	217	5	024300	024300 drosophila	624	242	2	09S1C9	09S1C9 porphyronom
552	6	2.0	217	9	094MQ3	094MQ3 bacterioph	625	242	4	09HBL2	09HBL2 homo sapien
553	6	2.0	217	10	09FHY7	09FHY7 arabidopsis	626	242	8	08WEN3	08WEN3 pogonomyme
554	6	2.0	217	17	08UIU2	08UIU2 pyrococcus	627	242	8	08W8K9	08W8K9 pogonomyme
555	6	2.0	218	8	09B522	09B522 lingua ung	628	242	17	096Z44	096Z44 sulfolobus
556	6	2.0	218	10	094925	094925 arabidopsis	629	243	8	08W7G3	08W7G3 pogonomyme
557	6	2.0	218	16	0932K2	0932K2 staphylococ	630	243	16	08XR40	08XR40 ralstonia s
558	6	2.0	218	17	027505	027505 methanobact	631	244	5	09KKF3	09KKF3 listeria mo
559	6	2.0	219	12	039325	039325 feline cali	632	244	5	09N650	09N650 manduca sex
560	6	2.0	219	12	056316	056316 feline cali	633	244	5	026300	026300 manduca sex
561	6	2.0	219	12	066865	066865 feline cali	634	244	10	098RR0	098RR0 guillardia
562	6	2.0	219	12	066868	066868 feline cali	635	244	12	091BB7	091BB7 spodoptera
563	6	2.0	219	12	066871	066871 feline cali	636	244	16	09KKK3	09KKK3 streptomyce
564	6	2.0	219	15	073115	073115 human immun	637	245	11	09D1X2	09D1X2 mus musculu
565	6	2.0	220	5	096701	096701 illyanasea o	638	245	16	092AL6	092AL6 listeria in
566	6	2.0	220	15	075086	075086 human immun	639	245	16	08Y6A3	08Y6A3 listeria mo
567	6	2.0	220	16	09PB28	09PB28 xylella fas	640	246	2	09RH35	09RH35 ehrlichia c
568	6	2.0	220	17	09YF60	09YF60 aeropyrum p	641	247	5	09VPEC2	09VPEC2 drosophila
569	6	2.0	221	10	09XJ12	09XJ12 oryza sativ	642	248	2	09VPEC2	09VPEC2 drosophila
570	6	2.0	221	11	091X14	091X14 rattus norv	643	248	5	019091	019091 caenorhabdi
571	6	2.0	221	16	08XV58	08XV58 ralstonia s	644	248	16	094538	094538 bacillus su
572	6	2.0	222	2	09S314	09S314 corynebacte	645	249	2	09AJT2	09AJT2 thauera aro
573	6	2.0	222	16	091248	091248 pseudomonas	646	249	5	062042	062042 caenorhabdi
574	6	2.0	222	16	08ZG23	08ZG23 versinia pe	647	249	16	09A4Y4	09A4Y4 caulobacter
575	6	2.0	223	16	08XTY6	08XTY6 ralstonia s	648	250	4	09NSV7	09NSV7 homo sapien
576	6	2.0	223	17	09HSN4	09HSN4 halobacteri	649	250	5	009521	009521 caenorhabdi
577	6	2.0	224	2	046369	046369 chlorobium	650	251	1	057432	057432 methanosarc
578	6	2.0	224	8	094YV7	094YV7 pylaiaella l	651	251	2	086270	086270 lactococcus
579	6	2.0	224	10	040503	040503 nicotiana t	652	251	5	09VAK9	09VAK9 drosophila
580	6	2.0	225	5	021197	021197 caenorhabdi	653	251	12	09ENR1	09ENR1 hepatitis a
581	6	2.0	226	10	09C9J7	09C9J7 arabidopsis	654	251	12	09ENR9	09ENR9 hepatitis a
582	6	2.0	227	11	09CX65	09CX65 mus musculu	655	251	12	09ENQ8	09ENQ8 hepatitis a
583	6	2.0	227	16	09KR08	09KR08 vibrio chol	656	251	12	09ENQ7	09ENQ7 hepatitis a
584	6	2.0	227	16	098EH7	098EH7 rhizobium l	657	251	12	09ENQ6	09ENQ6 hepatitis a
585	6	2.0	228	2	0937K6	0937K6 erwinia chr	658	251	12	09ENQ5	09ENQ5 hepatitis a
586	6	2.0	228	10	09LXY4	09LXY4 arabidopsis	659	251	12	09ENQ4	09ENQ4 hepatitis a
587	6	2.0	228	16	09RW00	09RW00 deinococcus	660	251	12	09ENQ3	09ENQ3 hepatitis a
588	6	2.0	228	16	09K767	09K767 bacillus ha	661	251	12	09ENQ3	09ENQ3 hepatitis a
589	6	2.0	228	16	08ZG24	08ZG24 versinia pe	662	251	12	09ENQ0	09ENQ0 hepatitis a
590	6	2.0	228	16	08UJ16	08UJ16 agrobacteri	663	251	12	09ENP9	09ENP9 hepatitis a
591	6	2.0	229	16	09CM68	09CM68 pasteurella	664	251	12	09ENP8	09ENP8 hepatitis a
592	6	2.0	230	10	09AXR7	09AXR7 gonyaulax p	665	251	12	09ENP7	09ENP7 hepatitis a
593	6	2.0	230	10	09SCA7	09SCA7 lycopersico	666	251	12	09ENP6	09ENP6 hepatitis a
594	6	2.0	230	10	09C6X3	09C6X3 arabidopsis	667	251	12	09ENP5	09ENP5 hepatitis a
595	6	2.0	230	16	097HM4	097HM4 clostridium	668	251	12	09ENP4	09ENP4 hepatitis a
596	6	2.0	230	16	0922P7	0922P7 rhizobium m	669	251	12	09ENP3	09ENP3 hepatitis a
597	6	2.0	230	16	08XFE1	08XFE1 salmonella	670	251	12	09ENP2	09ENP2 hepatitis a
598	6	2.0	231	10	093W01	093W01 arabidopsis	671	251	12	09ENP1	09ENP1 hepatitis a
599	6	2.0	231	10	09SJV8	09SJV8 arabidopsis	672	251	12	09ENP0	09ENP0 hepatitis a
600	6	2.0	231	16	08YZ81	08YZ81 anabaena sp	673	251	12	09ENN9	09ENN9 hepatitis a

674	6	2.0	251	12	Q9ENN7	Q9enn7 hepatitis a	747	6	2.0	272	15	Q90UR9	Q90ur9 equine infe
675	6	2.0	251	12	Q9ENN6	Q9enn6 hepatitis a	748	6	2.0	272	16	Q8UCJ5	Q8ucj5 agrobacteri
676	6	2.0	251	12	Q9ENN5	Q9enn5 hepatitis a	749	6	2.0	273	6	Q95MP9	Q95mp9 bos taurus
677	6	2.0	251	12	Q9ENNA	Q9enn4 hepatitis a	750	6	2.0	273	10	Q98QJ4	Q98qj4 brassica ca
678	6	2.0	251	12	Q9ENN2	Q9enn2 hepatitis a	751	6	2.0	273	15	Q90UR8	Q90ur8 equine infe
679	6	2.0	251	17	Q98396	Q28396 archaeoglob	752	6	2.0	273	15	Q90UR6	Q90ur6 equine infe
680	6	2.0	252	2	Q93099	Q53099 microcystis	753	6	2.0	273	15	Q90UR0	Q90ur0 equine infe
681	6	2.0	252	16	Q8XGA5	Q8xga5 salmonella	754	6	2.0	273	15	Q90UN6	Q90un6 equine infe
682	6	2.0	253	16	Q8R570	Q8r570 thermoanaer	755	6	2.0	273	15	Q90UN5	Q90un5 equine infe
683	6	2.0	254	11	Q9CQ99	Q9cq99 mus musculus	756	6	2.0	273	15	Q90UN3	Q90un3 equine infe
684	6	2.0	255	2	Q32774	Q32774 lactococcus	757	6	2.0	273	15	Q90UM8	Q90um8 equine infe
685	6	2.0	255	6	Q9MZA6	Q9mza6 equus caball	758	6	2.0	273	15	Q90UL4	Q90ul4 equine infe
686	6	2.0	255	6	Q9MZ48	Q9mz48 sus scrofa	759	6	2.0	273	16	Q9KU04	Q9ku04 vibrio chol
687	6	2.0	255	16	Q8YJ63	Q8yj63 brucella me	760	6	2.0	274	5	Q95X94	Q95x94 caenorhabdi
688	6	2.0	256	3	Q9Y842	Q9y842 metarhizium	761	6	2.0	275	5	Q9W2F1	Q9w2f1 drosophila
689	6	2.0	256	10	Q9FWJ4	Q9fwj4 arabidopsis	762	6	2.0	275	2	Q9ADV2	Q9adv2 ehrlichia c
690	6	2.0	256	10	Q9FYX7	Q9fyx7 glycine max	763	6	2.0	275	2	Q93DD4	Q93dd4 ehrlichia c
691	6	2.0	256	17	Q29814	Q29814 archaeoglob	764	6	2.0	275	15	Q90UT1	Q90ut1 equine infe
692	6	2.0	256	17	Q29814	Q29814 archaeoglob	765	6	2.0	275	15	Q90US8	Q90us8 equine infe
693	6	2.0	257	5	Q18200	Q18200 caenorhabdi	766	6	2.0	275	15	Q90US3	Q90us3 equine infe
694	6	2.0	257	5	Q95PN4	Q95pn4 paracentrot	767	6	2.0	275	15	Q90UP4	Q90up4 equine infe
695	6	2.0	257	9	Q9B0E1	Q9b0e1 staphylococ	768	6	2.0	275	15	Q90UP3	Q90up3 equine infe
696	6	2.0	257	16	Q9PPN8	Q9ppn8 ureaplasma	769	6	2.0	275	15	Q90UL8	Q90ul8 equine infe
697	6	2.0	257	16	Q98E38	Q98e38 rhizobium l	770	6	2.0	275	17	Q9HKC7	Q9hkc7 thermoplas
698	6	2.0	257	17	Q8YMX1	Q8ymx1 anabaena sp	771	6	2.0	276	2	Q85817	Q85817 ehrlichia c
699	6	2.0	257	17	P95890	P95890 sulfolobus	772	6	2.0	276	2	Q93DD1	Q93dd1 ehrlichia c
700	6	2.0	259	10	P93786	P93786 solanum tub	773	6	2.0	276	15	Q90US4	Q90us4 equine infe
701	6	2.0	259	10	Q94BG5	Q94bg5 frankenia c	774	6	2.0	276	15	Q90UP0	Q90up0 equine infe
702	6	2.0	260	16	Q9AAU9	Q9aa9 caulobacter	775	6	2.0	276	16	Q9RYH3	Q9ryh3 deinococcus
703	6	2.0	260	16	Q97GA7	Q97ga7 clostridium	776	6	2.0	277	15	Q90UJ6	Q90uj6 equine infe
704	6	2.0	261	4	Q96B80	Q96b80 homo sapien	777	6	2.0	277	15	Q90UU5	Q90uu5 equine infe
705	6	2.0	261	12	Q9IPD8	Q9ipd8 saeporo-lik	778	6	2.0	277	15	Q90UU4	Q90uu4 equine infe
706	6	2.0	261	16	Q8Y341	Q8y341 raietonia s	779	6	2.0	277	15	Q90UU3	Q90uu3 equine infe
707	6	2.0	262	5	Q76289	Q76289 heliocidari	780	6	2.0	277	15	Q90UU2	Q90uu2 equine infe
708	6	2.0	262	10	Q9LYI6	Q9lyi6 arabidopsis	781	6	2.0	277	15	Q90UU1	Q90uu1 equine infe
709	6	2.0	262	16	Q8RIE1	Q8rif1 fusobacteri	782	6	2.0	277	15	Q90UT9	Q90ut9 equine infe
710	6	2.0	263	2	Q84989	Q84989 rhodococcus	783	6	2.0	277	15	Q90UT8	Q90ut8 equine infe
711	6	2.0	263	2	Q82577	Q82577 zymomonas m	784	6	2.0	277	15	Q90UT3	Q90ut3 equine infe
712	6	2.0	263	2	Q05346	Q05346 synecococc	785	6	2.0	277	15	Q90UM3	Q90um3 equine infe
713	6	2.0	263	16	Q9PPK9	Q9ppk9 campylobact	786	6	2.0	277	15	Q90UL3	Q90ul3 equine infe
714	6	2.0	264	2	Q9A029	Q9a029 streptococ	787	6	2.0	277	15	Q90UL6	Q90ul6 equine infe
715	6	2.0	264	2	Q9ADW3	Q9adw3-ehlichia c	788	6	2.0	277	16	Q92782	Q92782 chlamydia p
716	6	2.0	264	16	Q25990	Q25990 helicobacte	789	6	2.0	277	16	Q24918	Q24918 helicobacte
717	6	2.0	264	16	Q98CV1	Q98cv1 rhizobium l	790	6	2.0	277	16	Q92MY0	Q92my0 helicobacte
718	6	2.0	265	10	Q987N6	Q987n6 arabidopsis	791	6	2.0	278	2	Q52299	Q52299 anabaena az
719	6	2.0	266	2	Q9JN23	Q9jnz3 micrococcus	792	6	2.0	278	10	Q00980	Q00980 luffa cylin
720	6	2.0	266	16	Q8X7G5	Q8x7g5 escherichia	793	6	2.0	278	15	Q90UR1	Q90ur1 equine infe
721	6	2.0	266	16	Q9EWM0	Q9ewm0 streptomyce	794	6	2.0	278	15	Q90UP8	Q90up8 equine infe
722	6	2.0	267	8	Q9T233	Q9t233 phytophthor	795	6	2.0	278	15	Q90UL6	Q90ul6 equine infe
723	6	2.0	268	5	Q20434	Q20434 caenorhabdi	796	6	2.0	278	16	Q51287	Q51287 borrelia bu
724	6	2.0	269	4	Q9BV00	Q9bv00 homo sapien	797	6	2.0	278	16	Q8Y8W3	Q8y8w3 anabaena sp
725	6	2.0	269	4	Q9BXC2	Q9bxc2 homo sapien	798	6	2.0	279	3	Q9US20	Q9us20 schizosacch
726	6	2.0	269	5	Q15781	Q15781 babesia bov	799	6	2.0	279	15	Q90UL1	Q90ul1 equine infe
727	6	2.0	269	5	Q8STE7	Q8ste7 drosophila	800	6	2.0	279	16	Q928M9	Q928m9 listeria in
728	6	2.0	269	5	Q85VP3	Q8avp3 encephalito	801	6	2.0	279	16	Q8Y4N2	Q8y4n2 listeria mo
729	6	2.0	269	10	Q9FVH9	Q9fvh9 prunus arme	802	6	2.0	280	10	Q9ZQD9	Q9zqd9 arabidopsis
730	6	2.0	269	16	P73457	P73457 synecocyst	803	6	2.0	280	10	Q40502	Q40502 nicotiana t
731	6	2.0	269	16	Q99Z72	Q99z72 streptococ	804	6	2.0	280	15	Q36343	Q36343 equine infe
732	6	2.0	270	10	Q8S6Z1	Q8s6z1 oryza sativ	805	6	2.0	280	15	Q90UP5	Q90up5 equine infe
733	6	2.0	270	16	Q8YMP6	Q8ym6 anabaena sp	806	6	2.0	280	15	Q90UL5	Q90ul5 equine infe
734	6	2.0	271	2	Q9F471	Q9f471 ehrlichia c	807	6	2.0	280	16	Q8RAG2	Q8rag2 thermoanaer
735	6	2.0	271	5	Q95MU7	Q95mu7 giardia lam	808	6	2.0	281	5	Q966N5	Q966n5 caenorhabdi
736	6	2.0	271	5	Q8T6U8	Q8t6u8 giardia lam	809	6	2.0	281	5	Q94E44	Q94e44 oryza sativ
737	6	2.0	272	11	Q39418	Q39418 beta vulgar	810	6	2.0	281	12	Q56871	Q56871 gallid herp
738	6	2.0	272	11	Q9D1C6	Q9dlc6 mus musculu	811	6	2.0	281	15	Q90UN2	Q90un2 equine infe
739	6	2.0	272	15	Q90UT2	Q90ut2 equine infe	812	6	2.0	281	15	Q90UN0	Q90un0 equine infe
740	6	2.0	272	15	Q90UT0	Q90ut0 equine infe	813	6	2.0	281	15	Q90UM0	Q90um0 equine infe
741	6	2.0	272	15	Q90US9	Q90us9 equine infe	814	6	2.0	281	15	Q90UL9	Q90ul9 equine infe
742	6	2.0	272	15	Q90US7	Q90us7 equine infe	815	6	2.0	281	17	Q8TIU0	Q8tiu0 methanosaer
743	6	2.0	272	15	Q90US6	Q90us6 equine infe	816	6	2.0	282	2	Q55068	Q55068 synecocyst
744	6	2.0	272	15	Q90US5	Q90us5 equine infe	817	6	2.0	282	2	Q9R443	Q9r443 cowdria rum
745	6	2.0	272	15	Q90US2	Q90us2 equine infe	818	6	2.0	282	2	Q9W441	Q9ww41 cowdria rum
746	6	2.0	272	15	Q90US1	Q90us1 equine infe	819	6	2.0	282	2	Q9AY79	Q9ay79 oryza sativ

820	6	2.0	282	15	Q90UR5	Q90ur5 equine infe	893	298	10	Q9LKV2	Q9lkv2 arabidopsis
821	6	2.0	282	15	Q90UR3	Q90ur3 equine infe	894	298	16	Q98IC0	Q98ic0 rhizobium l
822	6	2.0	282	15	Q90UR2	Q90ur2 equine infe	895	299	3	Q5908	Q5908 neurospora
823	6	2.0	282	15	Q90UQ7	Q90uq7 equine infe	896	299	16	Q9KZ12	Q9kz12 streptomyce
824	6	2.0	282	15	Q90UP2	Q90up2 equine infe	897	300	2	Q8RMD4	Q8rmd4 actinoplan
825	6	2.0	282	15	Q90UN4	Q90un4 equine infe	898	300	16	Q66604	Q66604 aquifex ao
826	6	2.0	282	15	Q90UM7	Q90um7 equine infe	899	300	16	Q66604	Q66604 aquifex ao
827	6	2.0	282	15	Q90UL2	Q90ul2 equine infe	900	300	16	Q8KZ40	Q8kz40 streptomyce
828	6	2.0	282	16	P73531	P73531 synechocyst	901	300	16	Q8XQC1	Q8xqc1 ralestonia s
829	6	2.0	282	16	P73976	P73976 synechocyst	902	301	5	Q8TT25	Q8tt25 methanosarc
830	6	2.0	282	16	Q55645	Q55645 synechocyst	903	301	5	Q9N480	Q9n480 caenorhabdi
831	6	2.0	282	16	Q57248	Q57248 synechocyst	904	301	11	Q8VGN4	Q8vgn4 mus musculu
832	6	2.0	282	16	P75027	P75027 synechocyst	905	301	16	Q8HUE0	Q8hue0 pseudomonas
833	6	2.0	283	2	O52103	O52103 ehrlichia c	906	301	16	Q97ML8	Q97ml8 clostridium
834	6	2.0	283	11	Q9CU44	Q9cu44 mus musculu	907	302	17	Q979G6	Q979g6 thermoplasm
835	6	2.0	283	12	Q9YV6	Q9ymv6 lymantria d	908	302	1	Q50755	Q50755 methanobact
836	6	2.0	283	15	Q36344	Q36344 equine infe	909	302	5	Q967P5	Q967p5 coleobothru
837	6	2.0	283	15	Q36347	Q36347 equine infe	910	302	16	Q9K3M4	Q9k3m4 streptomyce
838	6	2.0	283	15	Q90UQ3	Q90uq3 equine infe	911	302	17	Q27907	Q27907 methanobact
839	6	2.0	283	15	Q90UQ0	Q90uq0 equine infe	912	303	2	O68134	O68134 rhodobacter
840	6	2.0	283	15	Q90UN7	Q90un7 equine infe	913	303	16	Q8XKL6	Q8xkl6 clostridium
841	6	2.0	283	15	Q90UN1	Q90un1 equine infe	914	303	17	Q85704	Q85704 staphylococ
842	6	2.0	284	2	Q34174	Q34174 pseudomonas	915	304	16	Q9HRM8	Q9hrm8 halobacteri
843	6	2.0	284	15	Q36340	Q36340 equine infe	916	304	16	Q98R98	Q98r98 mycoplasma
844	6	2.0	284	15	Q36341	Q36341 equine infe	917	304	16	Q97LX7	Q97lx7 clostridium
845	6	2.0	284	15	Q36342	Q36342 equine infe	918	304	16	Q8YOC0	Q8yoc0 ralestonia s
846	6	2.0	284	15	Q36345	Q36345 equine infe	919	304	16	Q98RD2	Q98rd2 mycoplasma
847	6	2.0	284	15	Q36346	Q36346 equine infe	920	305	17	Q97Z11	Q97z11 sulfolobus
848	6	2.0	284	15	Q36348	Q36348 equine infe	921	306	4	Q95901	Q95901 homo sapien
849	6	2.0	284	15	Q36349	Q36349 equine infe	922	306	10	Q9LHY8	Q9lhy8 oryza sativ
850	6	2.0	284	15	Q90UM1	Q90um1 equine infe	923	306	10	Q8SOM3	Q8som3 oryza sativ
851	6	2.0	285	2	Q9L622	Q9l622 ehrlichia c	924	306	16	Q8XYM9	Q8xym9 ralestonia s
852	6	2.0	285	2	Q9ACJ3	Q9acj3 ehrlichia c	925	307	2	P94308	P94308 bacillus ps
853	6	2.0	285	2	Q45782	Q45782 bacillus th	926	307	2	Q8RJU8	Q8rju8 streptomyce
854	6	2.0	285	4	Q96B70	Q96bt0 homo sapien	927	307	12	Q84404	Q84404 peanut bud
855	6	2.0	285	10	Q9LMC7	Q9lmc7 arabidopsis	928	307	16	Q8YNU4	Q8ynu4 anabaena sp
856	6	2.0	285	16	Q9HZ92	Q9hz92 pseudomonas	929	307	16	Q8XYZ9	Q8xyz9 ralestonia s
857	6	2.0	287	16	Q9LIL3	Q9lil3 streptomyce	930	308	10	Q40668	Q40668 oryza sativ
858	6	2.0	287	17	O26472	O26472 methanobact	931	308	16	Q82P12	Q82p12 salmonella
859	6	2.0	288	2	Q9EU95	Q9eu95 escherichia	932	309	2	Q85774	Q85774 rhizobium l
860	6	2.0	288	2	Q9ETB2	Q9etb2 escherichia	933	309	16	Q926E6	Q926e6 rhizobium m
861	6	2.0	288	2	Q9F6H2	Q9f6h2 escherichia	934	310	4	Q9NZ35	Q9nz35 homo sapien
862	6	2.0	288	2	Q9F6H0	Q9f6h0 escherichia	935	310	4	Q96D57	Q96d57 homo sapien
863	6	2.0	288	2	Q9F6G9	Q9f6g9 escherichia	936	310	10	Q9ZNV7	Q9znv7 arabidopsis
864	6	2.0	288	2	Q9F6G0	Q9f6g0 escherichia	937	310	17	Q9YC70	Q9yc70 aeropyrum p
865	6	2.0	288	2	Q9F6F6	Q9f6f6 escherichia	938	311	2	Q93SF5	Q93sf5 eubacterium
866	6	2.0	288	2	Q9F6P4	Q9f6p4 escherichia	939	311	4	Q9BT20	Q9bt20 homo sapien
867	6	2.0	288	2	Q9ETR6	Q9etr6 escherichia	940	311	12	Q8QMS1	Q8qms1 cowpox viru
868	6	2.0	288	2	Q9ETA5	Q9eta5 escherichia	941	312	10	Q40309	Q40309 medicago sa
869	6	2.0	288	3	Q34330	Q34330 saccharomyc	942	312	10	Q40310	Q40310 medicago sa
870	6	2.0	288	10	Q8S870	Q8s870 oryza sativ	943	312	10	Q40333	Q40333 medicago sa
871	6	2.0	288	15	Q909J6	Q909j6 human immun	944	312	10	Q43555	Q43555 medicago sa
872	6	2.0	289	2	Q9F151	Q9f151 burkholderi	945	312	10	Q43556	Q43556 medicago sa
873	6	2.0	289	2	Q9LAH1	Q9lah1 acetobacter	946	312	16	Q9KEX4	Q9kex4 bacillus ha
874	6	2.0	289	5	O02153	O02153 caenorhabdi	947	312	16	Q8Y306	Q8y306 ralestonia s
875	6	2.0	289	16	Q8YUS1	Q8yus1 anabaena sp	948	313	2	Q55015	Q55015 streptomyce
876	6	2.0	289	16	Q8R7Q1	Q8r7q1 thermoaer	949	313	10	Q43610	Q43610 pyrus commu
877	6	2.0	290	15	Q90UP9	Q90up9 equine infe	950	313	11	Q9CZB7	Q9czb7 mus musculu
878	6	2.0	290	16	Q9FBZ9	Q9fbz9 streptomyce	951	313	11	Q9WTX2	Q9wtx2 mus musculu
879	6	2.0	291	3	Q9Y7C5	Q9y7c5 emericella	952	314	10	Q92U71	Q92u71 arabidopsis
880	6	2.0	292	5	Q9V3T7	Q9v3t7 drosophila	953	314	10	Q8S4C1	Q8s4c1 pueraria lo
881	6	2.0	293	16	Q9RKFS	Q9rkfs streptomyce	954	314	17	Q8TJ44	Q8tj44 methanosarc
882	6	2.0	294	16	Q9K792	Q9k792 bacillus ha	955	315	5	Q9NUY3	Q9ny3 leishmania
883	6	2.0	294	16	Q8R9X7	Q8r9x7 thermoaer	956	315	10	Q96426	Q96426 glycyrrhiza
884	6	2.0	295	5	Q8SWM1	Q8swm1 encephalito	957	315	16	Q8YFMI	Q8yfm1 brucella me
885	6	2.0	295	16	Q92F79	Q92f79 listeria in	958	316	2	O68147	O68147 escherichia
886	6	2.0	295	16	Q8YAE2	Q8yae2 listeria mo	959	316	2	O68150	O68150 escherichia
887	6	2.0	295	17	O26863	O26863 methanobact	960	316	2	O68152	O68152 escherichia
888	6	2.0	296	2	Q8RTS9	Q8rts9 uncultured	961	316	2	O68153	O68153 escherichia
889	6	2.0	296	16	Q99TT9	Q99tt9 staphylococ	962	316	2	O68154	O68154 escherichia
890	6	2.0	297	2	Q48482	Q48482 klebsiella	963	316	2	Q47700	Q47700 escherichia
891	6	2.0	297	2	Q937N2	Q937n2 photorhabdu	964	316	2	Q47701	Q47701 escherichia
892	6	2.0	298	2	Q9ACU6	Q9acj6 ehrlichia c	965	316	2	Q70099	Q70099 escherichia

966 Q9BQ66 homo sapien  
967 Q96425 glycyrrhiza  
968 Q92BC1 listeria in  
969 Q8R725 thermoanaer  
970 Q68148 escherichia  
971 Q68149 escherichia  
972 Q68151 escherichia  
973 Q9K1Q1 neisseria m  
974 Q9F3C2 streptomyce  
975 Q93C92 actinomadir  
976 Q9H8P0 homo sapien  
977 Q9T0R4 bacterioph  
978 Q37862 bacterioph  
979 Q6802 nicotiana t  
980 Q9DWS8 rat cytomeg  
981 Q8Y2P5 anabaena sp  
982 Q931F9 propionibac  
983 Q39774 glycyrrhiza  
984 Q9SU03 arabidopsis  
985 Q92T34 rhizobium m  
986 Q23259 caenorhabdi  
987 Q9C550 leishmania  
988 Q8VF14 mus musculu  
989 Q9JW1 neisseria m  
990 P96826 mycobacteri  
991 Q9K1W1 vibrio chol  
992 Q98G06 rhizobium l  
993 Q9N8B0 trypanosoma  
994 Q9ZP05 arabidopsis  
995 Q9CM30 pasteuralla  
996 Q9SU06 arabidopsis  
997 Q8TJ00 methanosarc  
998 Q985G2 escherichia  
999 Q06763 mycoplasma  
1000

## ALIGNMENTS

RESULT 1  
Q9F477 PRELIMINARY; PRT; 293 AA.  
ID Q9F477  
AC Q9F477  
DT 01-MAR-2001 (TremBLrel. 16, Created)  
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)  
DE P28-1 (Major outer membrane protein p30-5).  
GN P28-1 OR P30-5.  
OS Ehrlichia canis.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=944;  
RN [1]\_TaxID=944;  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAKE;  
RX MEDLINE=99242757; PubMed=10225842;  
RA McBride J.W., Yu, X.J., Walker D.H.;  
RT "Molecular cloning of the gene for a conserved major immunoreactive  
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic  
RT antigen."  
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAKE;  
RX MEDLINE=20432107; PubMed=10974556;  
RA McBride J.W., Yu X.J., Walker D.H.;  
RT "A conserved, transcriptionally active p28 multigene locus of  
RT Ehrlichia canis."  
RL Gene 254:245-252(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OKLAHOMA;  
RX MEDLINE=98371112; PubMed=9705412;  
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;  
RT "Cloning and characterization of multigenes encoding the  
RT immunodominant 30-kilodalton major outer membrane proteins of  
RT Ehrlichia canis and application of the recombinant protein for  
RT serodiagnosis."  
RL J. Clin. Microbiol. 36:2671-2680(1998).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OKLAHOMA;  
RX MEDLINE=21153566; PubMed=11254561;  
RA Ohashi N., Rikihisa Y., Unver A.;  
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer  
RT Membrane Protein Multigene Family in Ehrlichia canis and E.  
RT chaffeensis."  
RL Infect. Immun. 69:2083-2091(2001).  
DR EMBL; AF082744; AAG14356.1; -  
DR EMBL; AF078553; AAK28695.1; -  
DR InterPro; IPR002566; Surface\_Ag\_msp4.  
DR Pfam; PF01617; Surface\_Ag\_2; 1.0A15E94BLD2A3B65 CRC64;  
SQ SEQUENCE 293 AA; 32232 MW; 0A15E94BLD2A3B65 CRC64;  
Query Match 100.0%; Score 293; DB 2; Length 293;  
Best Local Similarity 100.0%; Pred. No. 1.2e-305;  
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MNKKLFTIINTVLVCLLSLPNISSKAINNNAKKYGLYISGQYKPSVSFVSFKET 60  
DB 1 MNKKLFTIINTVLVCLLSLPNISSKAINNNAKKYGLYISGQYKPSVSFVSFKET 60  
QY 61 NVITKNLIALKKVDSTETKTDSVGLSNPSNFTIPTAVFDQNSVNFNGTGYTFAEGT 120  
DB 61 NVITKNLIALKKVDSTETKTDSVGLSNPSNFTIPTAVFDQNSVNFNGTGYTFAEGT 120  
QY 121 RVEIEGSEYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPEKVSNSIFHTVMRDGLS 180  
DB 121 RVEIEGSEYEEFDVKNPGGYTLSDAYRYFALAREMKGNSFTPEKVSNSIFHTVMRDGLS 180  
QY 181 IISVIVNVVDFSLNNLSISPIYICGGAGVDAIEFDFVLHFKFAYQSKLGIAYSLPNLSL 240  
DB 181 IISVIVNVVDFSLNNLSISPIYICGGAGVDAIEFDFVLHFKFAYQSKLGIAYSLPNLSL 240  
QY 241 FASLYYHKVMGNQPKNLNVQHVLAELASIPKITSAVATLNICYFGGEIGARLTF 293  
DB 241 FASLYYHKVMGNQPKNLNVQHVLAELASIPKITSAVATLNICYFGGEIGARLTF 293  
RESULT 2  
Q9L6Y1 PRELIMINARY; PRT; 288 AA.  
ID Q9L6Y1  
AC Q9L6Y1  
DT 01-OCT-2000 (TremBLrel. 15, Created)  
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)  
DE P28-13 (Fragment).  
GN P28-13.  
OS Ehrlichia chaffeensis.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=945;  
RN [1]\_TaxID=945;  
RP SEQUENCE FROM N.A.  
RC STRAIN=ARKANSAS;  
RA Yu X.J., McBride J.W., Zhang X., Walker D.H.;  
RT "Characterization of the complete transcriptionally active Ehrlichia  
RT chaffeensis 28 kDa outer membrane protein multigene family."  
RL Gene 248:29-68(2000).  
DR EMBL; AF230642; AAF73422.1; -  
DR InterPro; IPR002566; Surface\_Ag\_msp4.  
DR Pfam; PF01617; Surface\_Ag\_2; 1.  
FT NON\_TER 288 288  
SQ SEQUENCE 288 AA; 31955 MW; 1ED203FBAED9CAEF CRC64;

```

Query Match      8.5%; Score 25; DB 2; Length 289;
Best Local Similarity 100.0%; Pred.No. 5.4e-18;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GLYISGQYKPSVSVFNSFNSVKETNV 62
    |||||
DB 42 GLYISGQYKPSVSVFNSFNSVKETNV 66

RESULT 3
ID O52102 PRELIMINARY; PRT; 297 AA.
AC O52102;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Major outer membrane protein OMP-1A.
GN OMP-1A.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP PARTIAL SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer Membrane Protein Multigene Family in Ehrlichia canis and E. chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U72291; AAC02935.2; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 297 AA; 32962 MW; 2ACA25B646BEE141 CRC64;

Query Match      8.5%; Score 25; DB 2; Length 297;
Best Local Similarity 100.0%; Pred.No. 5.5e-18;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GLYISGQYKPSVSVFNSFNSVKETNV 62
    |||||
DB 42 GLYISGQYKPSVSVFNSFNSVKETNV 66

RESULT 4
ID Q9ADV5 PRELIMINARY; PRT; 296 AA.
AC Q9ADV5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Major outer membrane protein P30-7.
GN P30-7.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the immunodominant 30-kilodalton major outer membrane proteins of Ehrlichia canis and application of the recombinant protein for serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer Membrane Protein Multigene Family in Ehrlichia canis and E. chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF078553; AAK28693.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 296 AA; 33281 MW; D001CAE71F748E53 CRC64;

Query Match      8.2%; Score 24; DB 2; Length 296;
Best Local Similarity 100.0%; Pred.No. 6.5e-17;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GLYISGQYKPSVSVFNSFNSVKETN 61
    |||||
DB 39 GLYISGQYKPSVSVFNSFNSVKETN 62

RESULT 5
ID Q9L6Y3 PRELIMINARY; PRT; 298 AA.
AC Q9L6Y3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P28-11 (Major outer membrane protein OMP-1H).
GN P28-11 OR OMP-1H.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer Membrane Protein Multigene Family in Ehrlichia canis and E. chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF230642; AAF73420.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 298 AA; 33026 MW; C2693B131D8CD2D5 CRC64;

Query Match      7.5%; Score 22; DB 2; Length 298;
Best Local Similarity 100.0%; Pred.No. 9.2e-15;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 YISQYKPSVSVFNSFNSVKETN 61
    |||||
DB 39 YISQYKPSVSVFNSFNSVKETN 60

```

```

RESULT 6
Q9L6Y2
ID Q9L6Y2 PRELIMINARY; PRT; 300 AA.
AC Q9L6Y2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P28-12 (Major outer membrane protein OMP-12).
GN P28-12 OR OMP-12.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RA Yu X.J., McBride J.W., Zhang X., Walker D.H.;
RT "Characterization of the complete transcriptionally active Ehrlichia
chaffeensis 28 kDa outer membrane protein multigene family.";
RL Gene 248:29-68(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
Membrane Protein Multigene Family in Ehrlichia canis and E.
chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF230642; AAF73419.1; -.
DR EMBL; U72291; AAK28672.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 300 AA; 33044 MW; 47B6A5A304C000B3 CRC64;

Query Match 5.8%; Score 17; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 2.2e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 VMGNQFKNLNVQHV AEL 265
DB 256 VMGNQFKNLNVQHV AEL 272

RESULT 7
Q9L6Y4
ID Q9L6Y4 PRELIMINARY; PRT; 291 AA.
AC Q9L6Y4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P28-10.
GN P28-10.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RA Yu X.J., McBride J.W., Zhang X., Walker D.H.;
RT "Characterization of the complete transcriptionally active Ehrlichia
chaffeensis 28 kDa outer membrane protein multigene family.";
RL Gene 248:29-68(2000).
DR EMBL; AF230642; AAF73419.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 291 AA; 31951 MW; 571FC9C795EB8CB4 CRC64;

Query Match 5.5%; Score 16; DB 2; Length 291;
Best Local Similarity 100.0%; Pred. No. 2.5e-08;

```

```

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GLYISGQYKPSVS VFS 53
DB 37 GLYISGQYKPSVS VFS 52

RESULT 8
Q9ACJ0
ID Q9ACJ0 PRELIMINARY; PRT; 291 AA.
AC Q9ACJ0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Major outer membrane protein OMP-18.
GN OMP-18.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
Membrane Protein Multigene Family in Ehrlichia canis and E.
chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; U72291; AAK28670.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 291 AA; 31974 MW; 3A35FB8BF6DFE2E6 CRC64;

Query Match 5.5%; Score 16; DB 2; Length 291;
Best Local Similarity 100.0%; Pred. No. 2.5e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GLYISGQYKPSVS VFS 53
DB 37 GLYISGQYKPSVS VFS 52

RESULT 9
Q9L6Y6
ID Q9L6Y6 PRELIMINARY; PRT; 275 AA.
AC Q9L6Y6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P28-8 (Major outer membrane protein OMP-1X).
GN P28-8 OR OMP-1X.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RA Yu X.J., McBride J.W., Zhang X., Walker D.H.;
RT "Characterization of the complete transcriptionally active Ehrlichia
chaffeensis 28 kDa outer membrane protein multigene family.";
RL Gene 248:29-68(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
Membrane Protein Multigene Family in Ehrlichia canis and E.
chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF230642; AAF73417.1; -.

```

DR EMBL; U72291; AAK28668.1; --  
 DR InterPro; IPR002566; Surface\_Ag\_msp4.  
 DR Pfam; PF01617; Surface\_Ag\_2; 1.  
 SQ SEQUENCE 275 AA; 30476 MW; 21B2A84CA8ACAE62 CRC64;

Query Match 5.1%; Score 15; DB 2; Length 275;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 274 AVATLNIYFGGEIG 288.

Db 256 AVATLNIYFGGEIG 270

RESULT 10

ID Q9ADV9 PRELIMINARY; PRT; 289 AA.  
 AC Q9ADV9;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
 DE Major outer membrane protein P30-12.  
 GN P30-12.  
 OS Ehrlichia canis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Anaplasmataceae; Ehrlichia.  
 OX NCBI\_TaxID=944;  
 RN RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OKLAHOMA;  
 RX MEDLINE=98371112; PubMed=9705412;  
 RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;  
 RT "Cloning and characterization of multigenes encoding the  
 immunodominant 30-kilodalton major outer membrane proteins of  
 Ehrlichia canis and application of the recombinant protein for  
 serodiagnosis";  
 RL J. Clin. Microbiol. 36:2671-2680 (1998).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OKLAHOMA;  
 RX MEDLINE=98371112; PubMed=9705412;  
 RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;  
 RT "Cloning and characterization of multigenes encoding the  
 immunodominant 30-kilodalton major outer membrane proteins of  
 Ehrlichia canis and application of the recombinant protein for  
 serodiagnosis";  
 RL J. Clin. Microbiol. 36:2671-2680 (1998).  
 [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=OKLAHOMA;  
 RX MEDLINE=21153566; PubMed=11254561;  
 RA Ohashi N., Rikihisa Y., Unver A.;  
 RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer  
 Membrane Protein Multigene Family in Ehrlichia canis and E.  
 chaffeensis";  
 RL Infect. Immun. 69:2083-2091 (2001).  
 DR EMBL; AF078553; AAK28689.1; --  
 DR InterPro; IPR002566; Surface\_Ag\_msp4.  
 DR Pfam; PF01617; Surface\_Ag\_2; 1.  
 SQ SEQUENCE 289 AA; 32460 MW; BBABDA87935D424A CRC64;

Query Match 5.1%; Score 15; DB 2; Length 289;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 EIEGSYEEFDVKNPG 137

Db 122 EIEGSYEEFDVKNPG 136

RESULT 11

ID Q9ADV7 PRELIMINARY; PRT; 281 AA.  
 AC Q9ADV7;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
 DE Major outer membrane protein P30-9.  
 GN P30-9.  
 OS Ehrlichia canis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Anaplasmataceae; Ehrlichia.  
 OX NCBI\_TaxID=944;

RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=OKLAHOMA;  
 RX MEDLINE=98371112; PubMed=9705412;  
 RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;  
 RT "Cloning and characterization of multigenes encoding the  
 immunodominant 30-kilodalton major outer membrane proteins of  
 Ehrlichia canis and application of the recombinant protein for  
 serodiagnosis";  
 RL J. Clin. Microbiol. 36:2671-2680 (1998).  
 [2]

RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OKLAHOMA;  
 RX MEDLINE=21153566; PubMed=11254561;  
 RA Ohashi N., Rikihisa Y., Unver A.;  
 RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer  
 Membrane Protein Multigene Family in Ehrlichia canis and E.  
 chaffeensis";  
 RL Infect. Immun. 69:2083-2091 (2001).  
 DR EMBL; AF078553; AAK28691.1; --  
 DR InterPro; IPR002566; Surface\_Ag\_msp4.  
 DR Pfam; PF01617; Surface\_Ag\_2; 1.  
 SQ SEQUENCE 281 AA; 31525 MW; DE94CF01CD24D967 CRC64;

Query Match 4.4%; Score 13; DB 2; Length 281;  
 Best Local Similarity 100.0%; Pred. No. 4e-05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 272 TSAVATLNIYFG 284

Db 260 TSAVATLNIYFG 272

RESULT 12

Q9L6Y9  
 ID Q9L6Y9 PRELIMINARY; PRT; 279 AA.  
 AC Q9L6Y9;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE P28-6.  
 GN P28-6.  
 OS Ehrlichia chaffeensis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Anaplasmataceae; Ehrlichia.  
 OX NCBI\_TaxID=945;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ARKANSAS;  
 RA Yu X.J., McBride J.W., Zhang X., Walker D.H.;  
 RT "Characterization of the complete transcriptionally active Ehrlichia  
 chaffeensis 28 kDa outer membrane protein multigene family";  
 RL Gene 248:29-68 (2000).  
 DR EMBL; AF230842; AAF73414.1; --  
 DR InterPro; IPR002566; Surface\_Ag\_msp4.  
 DR Pfam; PF01617; Surface\_Ag\_2; 1.  
 SQ SEQUENCE 279 AA; 30827 MW; 83633D35EF61206B CRC64;

Query Match 3.8%; Score 11; DB 2; Length 279;  
 Best Local Similarity 100.0%; Pred. No. 0.0056;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 52 FSNFSVKETNV 62

Db 53 FSNFSVKETNV 63

RESULT 13

ID Q9ACJ2 PRELIMINARY; PRT; 279 AA.  
 AC Q9ACJ2;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)  
DE Major outer membrane protein OMP-IV.  
GN OMP-IV.  
OS Ehrlichia chaffeensis.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=945;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ARKANSAS;  
RX MEDLINE=21153566; PubMed=11254561;  
RA Ohashi N., Rikihisa Y., Unver A.;  
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer  
RT Membrane Protein Multigene Family in Ehrlichia canis and E.  
RT chaffeensis.";  
RL Infect. Immun. 69:2083-2091(2001).  
DR EMBL; U72291; AAK28665.1; -;  
DR InterPro; IPR002566; Surface\_Ag\_msp4.  
DR Pfam; PF01617; Surface\_Ag\_2; 1.  
SQ SEQUENCE 279 AA; 30843 MW; 83633D35F2A0FC6B CRC64;

Query Match 3.8%; Score 11; DB 2; Length:279;  
Best Local Similarity 100.0%; Pred. No. 0.0056;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 PSNFSVKETNV 62

|||||

DB 53 PSNFSVKETNV 63

RESULT 14  
Q9L6V7  
ID Q9L6V7 PRELIMINARY; PRT; 283 AA.  
AC Q9L6V7;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE P28-7 (Major outer membrane protein OMP-IV).  
GN P28-7 OR OMP-IV.  
OS Ehrlichia chaffeensis.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=945;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ARKANSAS;  
RX MEDLINE=21153566; PubMed=11254561;  
RA Ohashi N., Rikihisa Y., Unver A.;  
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer  
RT Membrane Protein Multigene Family in Ehrlichia canis and E.  
RT chaffeensis.";  
RL Infect. Immun. 69:2083-2091(2001).  
DR EMBL; U72291; AAK28665.1; -;  
DR InterPro; IPR002566; Surface\_Ag\_msp4.  
DR Pfam; PF01617; Surface\_Ag\_2; 1.  
SQ SEQUENCE 283 AA; 31662 MW; 8F4778B09F1CD96E CRC64;

Query Match 3.8%; Score 11; DB 2; Length:283;  
Best Local Similarity 100.0%; Pred. No. 0.0056;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 SYEEFDVKNPG 137

|||||

DB 122 SYEEFDVKNPG 132

RESULT 15  
Q9L6Z0  
ID Q9L6Z0 PRELIMINARY; PRT; 295 AA.  
AC Q9L6Z0;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE P28-5 (Major outer membrane protein OMP-1U).  
GN P28-5 OR OMP-1U.  
OS Ehrlichia chaffeensis.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=945;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ARKANSAS;  
RX MEDLINE=21153566; PubMed=11254561;  
RA Yu X.J., McBride J.W., Zhang X., Walker D.H.;  
RT "Characterization of the complete transcriptionally active Ehrlichia  
RT chaffeensis 28 kDa outer membrane protein multigene family.";  
RL Gene 248:29-68(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ARKANSAS;  
RX MEDLINE=21153566; PubMed=11254561;  
RA Ohashi N., Rikihisa Y., Unver A.;  
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer  
RT Membrane Protein Multigene Family in Ehrlichia canis and E.  
RT chaffeensis.";  
RL Infect. Immun. 69:2083-2091(2001).  
DR EMBL; AF230642; AAF73413.1; -;  
DR EMBL; U72291; AAK28664.1; -;  
DR InterPro; IPR002566; Surface\_Ag\_msp4.  
DR Pfam; PF01617; Surface\_Ag\_2; 1.  
SQ SEQUENCE 295 AA; 33616 MW; 4F67BD9B99E658FE CRC64;

Query Match 3.8%; Score 11; DB 2; Length:295;  
Best Local Similarity 100.0%; Pred. No. 0.0059;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GLYISGOYKPS 48

|||||

DB 38 GLYISGOYKPS 48

RESULT 16  
Q9ADV6  
ID Q9ADV6 PRELIMINARY; PRT; 299 AA.  
AC Q9ADV6;  
DT 01-JUN-2001 (TReMBLrel. 17, Created)  
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)  
DE Major outer membrane protein P30-8.  
GN P30-8.  
OS Ehrlichia canis.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=944;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OKLAHOMA;  
RX MEDLINE=98371112; PubMed=9705412;  
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;  
RT "Cloning and characterization of multigenes encoding the  
RT immunodominant 30-kilodalton major outer membrane proteins of  
RT Ehrlichia canis and application of the recombinant protein for  
RT serodiagnosis.";  
RL J. Clin. Microbiol. 36:2671-2680(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OKLAHOMA;  
RX MEDLINE=21153566; PubMed=11254561;  
RA Ohashi N., Rikihisa Y., Unver A.;



RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer  
 RT Membrane Protein Multigene Family in Ehrlichia canis and E.  
 RL Chaffeensis.";  
 RL Infect. Immun. 69:2083-2091(2001).  
 DR EMBL; AF078553; AAK28692.1; -.  
 DR InterPro; IPR002566; Surface\_Ag\_msp4.  
 DR Pfam; PF01617; Surface\_Ag\_2; 1.  
 SQ SEQUENCE 299 AA; 33085 MW; 14A29F3A1A3CC42D CRC64;

Query Match 3.4%; Score 11; DB 2; Length 289;  
 Best Local Similarity 100.0%; Pred. No. 0.0059;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 269 PKITSATVTLN 279  
 Db 275 PKITSATVTLN 285  
 .|.|.|.|.|.|.|.|.|.|.  
 .|.|.|.|.|.|.|.|.|.|.

## RESULT 17

Q9L6Y5 ID Q9L6Y5 PRELIMINARY; PRT; 285 AA.

AC Q9L6Y5;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE P28-9 (Major outer membrane protein OMP-1Y).  
 GN P28-9 OR OMP-1Y.  
 OS Ehrlichia chaffeensis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Anaplasmataceae; Ehrlichia.  
 OX NCBI\_TaxID=945;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ARKANSAS;  
 RA Yu X.J., McBride J.W., Zhang X., Walker D.H.;  
 RT "Characterization of the complete transcriptionally active Ehrlichia  
 RT chaffeensis 28 kDa outer membrane protein multigene family.";  
 RL Gene 248:29-68(2000).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=ARKANSAS;  
 RX MEDLINE=21153566; PubMed=11254561;  
 RA Ohashi N., Rikihisa Y., Unver A.;  
 RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer  
 RT Membrane Protein Multigene Family in Ehrlichia canis and E.  
 RT chaffeensis.";  
 RL Infect. Immun. 69:2083-2091(2001).  
 DR EMBL; AF230642; AAF73418.1; -.  
 DR EMBL; U72291; AAK28669.1; -.  
 DR InterPro; IPR002566; Surface\_Ag\_msp4.  
 DR Pfam; PF01617; Surface\_Ag\_2; 1.  
 SQ SEQUENCE 285 AA; 31711 MW; 576009467A42985D CRC64;

Query Match 3.4%; Score 10; DB 2; Length 285;  
 Best Local Similarity 100.0%; Pred. No. 0.067;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 SYEEFDVKNP 136  
 Db 123 SYEEFDVKNP 132  
 .|.|.|.|.|.|.|.|.|.|.  
 .|.|.|.|.|.|.|.|.|.|.

## RESULT 18

Q9ADV4 ID Q9ADV4 PRELIMINARY; PRT; 294 AA.

AC Q9ADV4;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Major outer membrane protein P30-6.  
 GN P30-6.  
 OS Ehrlichia canis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Anaplasmataceae; Ehrlichia.  
 OX NCBI\_TaxID=944;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OKLAHOMA;  
 RX MEDLINE=98371112; PubMed=9705412;  
 RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;  
 RT "Cloning and characterization of multigenes encoding the  
 RT immunodominant 30-kilodalton major outer membrane proteins of  
 RT Ehrlichia canis and application of the recombinant protein for  
 RT serodiagnosis.";  
 RL J. Clin. Microbiol. 36:2671-2680(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OKLAHOMA;  
 RX MEDLINE=21153566; PubMed=11254561;  
 RA Ohashi N., Rikihisa Y., Unver A.;  
 RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer  
 RT Membrane Protein Multigene Family in Ehrlichia canis and E.  
 RT chaffeensis.";  
 RL Infect. Immun. 69:2083-2091(2001).  
 DR EMBL; AF078553; AAK28694.1; -.  
 DR InterPro; IPR000425; MIP family.  
 DR InterPro; IPR002566; Surface\_Ag\_msp4.  
 DR Pfam; PF01617; Surface\_Ag\_2; 1.  
 DR PROSITE; PS00221; MIP; UNKNOWN 1.  
 SQ SEQUENCE 294 AA; 32869 MW; 570FAC76A5820867 CRC64;

Query Match 3.4%; Score 10; DB 2; Length 294;  
 Best Local Similarity 100.0%; Pred. No. 0.069;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 PSVSVSFNFS 56  
 Db 48 PSVSVSFNFS 57  
 .|.|.|.|.|.|.|.|.|.|.  
 .|.|.|.|.|.|.|.|.|.|.

## RESULT 19

Q9ADW1 ID Q9ADW1 PRELIMINARY; PRT; 278 AA.

AC Q9ADW1;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
 DE Major outer membrane protein P30-13.  
 GN P30-13.  
 OS Ehrlichia canis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Anaplasmataceae; Ehrlichia.  
 OX NCBI\_TaxID=944;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OKLAHOMA;  
 RX MEDLINE=98371112; PubMed=9705412;  
 RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;  
 RT "Cloning and characterization of multigenes encoding the  
 RT immunodominant 30-kilodalton major outer membrane proteins of  
 RT Ehrlichia canis and application of the recombinant protein for  
 RT serodiagnosis.";  
 RL J. Clin. Microbiol. 36:2671-2680(1998).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=OKLAHOMA;  
 RX MEDLINE=21153566; PubMed=11254561;  
 RA Ohashi N., Rikihisa Y., Unver A.;  
 RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer  
 RT Membrane Protein Multigene Family in Ehrlichia canis and E.  
 RT chaffeensis.";  
 RL Infect. Immun. 69:2083-2091(2001).  
 DR EMBL; AF078553; AAK28687.1; -.  
 DR InterPro; IPR002566; Surface\_Ag\_msp4.  
 DR Pfam; PF01617; Surface\_Ag\_2; 1.  
 SQ SEQUENCE 278 AA; 30829 MW; 3D2F10AD6412C64C CRC64;

```
Query Match          3.1%; Score 9; DB 2; Length 278;
Best Local Similarity 100.0%; Pred. No. 0.78;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 251 GNQFKNLNV 259
DB 236 GNQFKNLNV 244
|||||

RESULT 20
Q9ADV8 PRELIMINARY; PRT; 279 AA.
AC Q9ADV8;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE Major outer membrane protein P30-11.
GN P30-11.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL: AF078553; AAK28690.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 279 AA; 31379 MW; 07B356825E071E80 CRC64;

Query Match          3.1%; Score 9; DB 2; Length 279;
Best Local Similarity 100.0%; Pred. No. 0.78;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 LYISQYKRP 47
DB 38 LYISQYKRP 46
|||||

RESULT 21
Q8UOV7 PRELIMINARY; PRT; 98 AA.
AC Q8UOV7;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein PF1474.
GN PF1474.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
```

```
"The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL: AB010249; AAU81598.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 98 AA; 11004 MW; AA3EB361322B77AE CRC64;

Query Match          2.7%; Score 8; DB 17; Length 98;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 ITKNLIAL 70
DB 19 ITKNLIAL 26
|||||

RESULT 22
Q9AF99 PRELIMINARY; PRT; 265 AA.
AC Q9AF99;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE Major antigenic protein 1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UMPALA;
RA Bensaïd A., Maillard J.-C., Chantal I., Perez J.-M., Martinez D.;
RT "Cowdria ruminantium major antigenic protein 1 (map1) gene variants
RT are not geographically constrained.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL: AF355202; AAK27218.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 1
FT NON_TER 265
SQ SEQUENCE 265 AA; 28388 MW; 1EA6DB25CCBF634 CRC64;

Query Match          2.7%; Score 8; DB 2; Length 265;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 FDKVNPQG 138
DB 111 FDKVNPQG 118
|||||

RESULT 23
Q93E63 PRELIMINARY; PRT; 268 AA.
AC Q93E63;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Major antigenic protein MAP1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KISWANI;
RX MEDLINE=21539003; PubMed=11682561;
RA Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaïd A.,
RA Haydon D.T., van Heerden H., Allsopp B.A.;
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
RT Are Not Geographically Constrained and Show No Evidence of Having
RT Evolved under Positive Selection Pressure.";
RL J. Clin. Microbiol. 39:4200-4203(2001).
DR EMBL: AF368003; AAK98144.1; -.
```

```

DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 268
FT NON_TER 268
SQ SEQUENCE 268 AA; 28999 MW; F5306B6C741AADE0 CRC64;

Query Match      2.7%; Score 8; DB 2; Length 268;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 FDVKNPGG 138
Db 118 FDVKNPGG 125

RESULT 24
Q93E61 ID Q93E61 PRELIMINARY; PRT; 268 AA.
AC Q93E61
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major antigenic protein MAP1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LUDLOW;
RX MEDLINE=21539003; PubMed=11682561;
RA Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaid A.,
RA Haydon D.T., van Heerden H., Allsopp B.A.;
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
RT Are Not Geographically Constrained and Show No Evidence of Having
RT Evolved under Positive Selection Pressure.";
RL J. Clin. Microbiol. 39:4200-4203 (2001).
DR EMBL; AF368005; AAK98145.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 268
FT NON_TER 268
SQ SEQUENCE 268 AA; 29013 MW; F53071737405B2E0 CRC64;

Query Match      2.7%; Score 8; DB 2; Length 268;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 FDVKNPGG 138
Db 118 FDVKNPGG 125

RESULT 25
Q93E62 ID Q93E62 PRELIMINARY; PRT; 269 AA.
AC Q93E62
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major antigenic protein MAP1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KWANYANGA;
RX MEDLINE=21539003; PubMed=11682561;
RA Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaid A.,
RA Haydon D.T., van Heerden H., Allsopp B.A.;
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
RT Are Not Geographically Constrained and Show No Evidence of Having

```

```

RT Evolved under Positive Selection Pressure.";
RL J. Clin. Microbiol. 39:4200-4203 (2001).
DR EMBL; AF368004; AAK98145.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 269
FT NON_TER 269
SQ SEQUENCE 269 AA; 28933 MW; BFA72C857C29D95C CRC64;

Query Match      2.7%; Score 8; DB 2; Length 269;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 FDVKNPGG 138
Db 118 FDVKNPGG 125

RESULT 26
Q9AF98 ID Q9AF98 PRELIMINARY; PRT; 270 AA.
AC Q9AF98
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Major antigenic protein 1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CAMEROUN;
RA Bensaid A., Maillard J.-C., Chantal I., Perez J.-M., Martinez D.;
RT "Cowdria ruminantium major antigenic protein 1 (map1) gene variants
RT are not geographically constrained.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF355203; AAK27219.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 270
FT NON_TER 270
SQ SEQUENCE 270 AA; 29380 MW; 99BAAAFB9E48895A6 CRC64;

Query Match      2.7%; Score 8; DB 2; Length 270;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 FDVKNPGG 138
Db 115 FDVKNPGG 122

RESULT 27
Q9AF98 ID Q9AF98 PRELIMINARY; PRT; 271 AA.
AC Q9AF98
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Major antigenic protein 1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LUTALE;
RA Bensaid A., Maillard J.-C., Chantal I., Perez J.-M., Martinez D.;
RT "Cowdria ruminantium major antigenic protein 1 (map1) gene variants
RT are not geographically constrained.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF355201; AAK27217.1; -.

```

DR InterPro: IPR002566; Surface\_Ag\_msp4.  
DR Pfam: PF01617; Surface\_Ag\_2; 1.  
FT NON\_TER 1  
SQ SEQUENCE 271 AA; 29072 MW; 3829FDA9BE25645 CRC64;

Query Match 2.7%; Score 8; DB 2; Length 271;  
Best Local Similarity 100.0%; Pred. No. 9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 FDKVKNPGG 138  
|||||  
Db 116 FDKVKNPGG 123

## RESULT 28

Q9AMF6 PRELIMINARY; PRT; 272 AA.  
AC Q9AMF6;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Major antigenic protein MAP1 (Fragment).  
GN MAP1.  
OS Ehrlichia sp. 'South African canine'.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=152574;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Allsopp M.T., Allsopp B.A.;  
RT "A novel Ehrlichia detected in dogs in South Africa."  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF325176; AAK14320.1; -;  
DR InterPro: IPR002566; Surface\_Ag\_msp4.  
DR Pfam: PF01617; Surface\_Ag\_2; 1.  
FT NON\_TER 1  
SQ SEQUENCE 272 AA; 29225 MW; 43CEB95DEDD55BDE CRC64;

## Query Match

Best Local Similarity 2.7%; Score 8; DB 2; Length 272;  
Best Local Similarity 100.0%; Pred. No. 9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 FDKVKNPGG 138  
|||||  
Db 120 FDKVKNPGG 127

## RESULT 29

Q93E54 PRELIMINARY; PRT; 272 AA.  
AC Q93E54;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-WAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Major antigenic protein MAP1 (Fragment).  
OS Cowdria ruminantium.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=779;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=POKOASE;  
RX MEDLINE=21539003; PubMed=11682561;  
RA Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaïd A.,  
RA Haydon D.T., van Heerden H., Allsopp B.A.;  
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants  
RT Are Not Geographically Constrained and Show No Evidence of Having  
RT Evolved under Positive Selection Pressure."  
RL J. Clin. Microbiol. 39:4200-4203 (2001).  
DR EMBL; AF368013; AAK98153.1; -;  
DR InterPro: IPR002566; Surface\_Ag\_msp4.

DR Pfam: PF01617; Surface\_Ag\_2; 1.  
FT NON\_TER 1  
SQ SEQUENCE 272 AA; 29225 MW; 43CEB95DEDD55BDE CRC64;

Query Match 2.7%; Score 8; DB 2; Length 272;  
Best Local Similarity 100.0%; Pred. No. 9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 FDKVKNPGG 138  
|||||  
Db 120 FDKVKNPGG 127

## RESULT 30

Q9L6Z1 PRELIMINARY; PRT; 272 AA.  
AC Q9L6Z1;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE P28-4 (Major outer membrane protein OMP-17).  
GN P28-4 OR OMP-17.  
OS Ehrlichia chaffeensis.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=945;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ARKANSAS;  
RA Yu X.J., McBride J.W., Zhang X., Walker D.H.;  
RT "Characterization of the complete transcriptionally active Ehrlichia  
RT chaffeensis 28 Kda outer membrane protein multigene family."  
RL Gene 248:29-68 (2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ARKANSAS;  
RX MEDLINE=21153566; PubMed=11254561;  
RA Chashi N., Rikihisa Y., Unver A.;  
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer  
RT Membrane Protein Multigene Family in Ehrlichia canis and E.  
RT chaffeensis."  
RL Infect. Immun. 69:2083-2091 (2001).  
DR EMBL; AF230642; AAF73412.1; -;  
DR EMBL; U72291; AAK28663.1; -;  
DR InterPro: IPR002566; Surface\_Ag\_msp4.  
DR Pfam: PF01617; Surface\_Ag\_2; 1.  
SQ SEQUENCE 272 AA; 31234 MW; 40A6A805368B9D2F CRC64;

Query Match 2.7%; Score 8; DB 2; Length 272;  
Best Local Similarity 100.0%; Pred. No. 9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 245 YYHKVMGN 252  
|||||  
Db 223 YYHKVMGN 230

## RESULT 31

Q93E59 PRELIMINARY; PRT; 275 AA.  
AC Q93E59;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Major antigenic protein MAP1 (Fragment).  
OS Cowdria ruminantium.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Anaplasmataceae; Ehrlichia.  
OX NCBI\_TaxID=779;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MARA87/7;

```

RX MEDLINE=21539003; PubMed=11682561;
RA Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaid A.,
RA Haydon D.T., van Heerden H., Allsopp B.A.;
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
RT Are Not Geographically Constrained and Show No Evidence of Having
RT Evolved under Positive Selection Pressure.";
RL J. Clin. Microbiol. 39:4200-4203(2001).
DR EMBL; AF368008; AAK98148.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 1
FT NON_TER 275
SQ SEQUENCE 275 AA; 29722 MW; 71170E2A159B28DB CRC64;

Query Match 2.7%; Score 8; DB 2; Length 275;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 FDVKNPGG 138
Db 123 FDVKNPGG 130

RESULT 32
Q93E60 PRELIMINARY; PRT; 276 AA.
ID Q93E60
AC Q93E60;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major antigenic protein MAP1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MALI;
RX MEDLINE=21539003; PubMed=11682561;
RA Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaid A.,
RA Haydon D.T., van Heerden H., Allsopp B.A.;
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
RT Are Not Geographically Constrained and Show No Evidence of Having
RT Evolved under Positive Selection Pressure.";
RL J. Clin. Microbiol. 39:4200-4203(2001).
DR EMBL; AF368007; AAK98147.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 1
FT NON_TER 276
SQ SEQUENCE 276 AA; 29856 MW; F685387E1602A50E CRC64;

Query Match 2.7%; Score 8; DB 2; Length 276;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 FDVKNPGG 138
Db 122 FDVKNPGG 129

RESULT 33
Q93E53 PRELIMINARY; PRT; 276 AA.
ID Q93E53
AC Q93E53;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major antigenic protein MAP1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BLAUKRANS;
RX MEDLINE=21539003; PubMed=11682561;
RA Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaid A.,
RA Haydon D.T., van Heerden H., Allsopp B.A.;
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
RT Are Not Geographically Constrained and Show No Evidence of Having
RT Evolved under Positive Selection Pressure.";
RL J. Clin. Microbiol. 39:4200-4203(2001).
DR EMBL; AF368007; AAK98147.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 1
FT NON_TER 276
SQ SEQUENCE 276 AA; 29856 MW; F685387E1602A50E CRC64;

Query Match 2.7%; Score 8; DB 2; Length 276;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 FDVKNPGG 138
Db 122 FDVKNPGG 129

RESULT 34
Q93E65 PRELIMINARY; PRT; 277 AA.
ID Q93E65
AC Q93E65;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major antigenic protein MAP1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BLAUKRANS;
RX MEDLINE=21539003; PubMed=11682561;
RA Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaid A.,
RA Haydon D.T., van Heerden H., Allsopp B.A.;
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
RT Are Not Geographically Constrained and Show No Evidence of Having
RT Evolved under Positive Selection Pressure.";
RL J. Clin. Microbiol. 39:4200-4203(2001).
DR EMBL; AF368000; AAK98142.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 1
FT NON_TER 277
SQ SEQUENCE 277 AA; 29916 MW; 167B95113194237A CRC64;

Query Match 2.7%; Score 8; DB 2; Length 277;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 FDVKNPGG 138
Db 122 FDVKNPGG 129

RESULT 35
Q93E57 PRELIMINARY; PRT; 278 AA.
ID Q93E57
AC Q93E57;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major antigenic protein MAP1 (Fragment).
OS Cowdria ruminantium.

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SANKAT;
RX MEDLINE=21539003; PubMed=11682561;
RA Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaid A.,
RA Haydon D.T., van Heerden H., Allsopp B.A.;
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
RT Are Not Geographically Constrained and Show No Evidence of Having
RT Evolved under Positive Selection Pressure.";
RL J. Clin. Microbiol. 39:4200-4203(2001).
DR EMBL; AF368014; AAK98154.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 1
FT NON_TER 276
SQ SEQUENCE 276 AA; 29870 MW; E415387E1602A50E CRC64;

Query Match 2.7%; Score 8; DB 2; Length 276;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 FDVKNPGG 138
Db 122 FDVKNPGG 129

RESULT 34
Q93E65 PRELIMINARY; PRT; 277 AA.
ID Q93E65
AC Q93E65;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major antigenic protein MAP1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BLAUKRANS;
RX MEDLINE=21539003; PubMed=11682561;
RA Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaid A.,
RA Haydon D.T., van Heerden H., Allsopp B.A.;
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
RT Are Not Geographically Constrained and Show No Evidence of Having
RT Evolved under Positive Selection Pressure.";
RL J. Clin. Microbiol. 39:4200-4203(2001).
DR EMBL; AF368000; AAK98142.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 1
FT NON_TER 277
SQ SEQUENCE 277 AA; 29916 MW; 167B95113194237A CRC64;

Query Match 2.7%; Score 8; DB 2; Length 277;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 FDVKNPGG 138
Db 122 FDVKNPGG 129

RESULT 35
Q93E57 PRELIMINARY; PRT; 278 AA.
ID Q93E57
AC Q93E57;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major antigenic protein MAP1 (Fragment).
OS Cowdria ruminantium.

```

```

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
RN NCBI_TaxID=779;
RP SEQUENCE FROM N.A.
RX STRAIN=MORGESWAG2;
RX MEDLINE=21539003; PubMed=11682561;
RA Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaïd A.,
RA Haydon D.T., van Heerden H., Allsopp B.A.;
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
RT Are Not Geographically Constrained and Show No Evidence of Having
RT Evolved under Positive Selection Pressure.";
RL J. Clin. Microbiol. 39:4200-4203 (2001).
DR EMBL; AF368010; AAK98150.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 1
FT NON_TER 278
SQ SEQUENCE 278 AA; 30048 MW; 6C74669F156D4980 CRC64;

Query Match 2.7%; Score 8; DB 2; Length 278;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 FDVKNPGG 138
DB 123 FDVKNPGG 130

RESULT 36
Q33E56 PRELIMINARY; PRT; 278 AA.
AC Q33E56;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major antigenic protein MAPI (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RP SEQUENCE FROM N.A.
RX STRAIN=NONILE;
RX MEDLINE=21539003; PubMed=11682561;
RA Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaïd A.,
RA Haydon D.T., van Heerden H., Allsopp B.A.;
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
RT Are Not Geographically Constrained and Show No Evidence of Having
RT Evolved under Positive Selection Pressure.";
RL J. Clin. Microbiol. 39:4200-4203 (2001).
DR EMBL; AF368011; AAK98151.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 1
FT NON_TER 278
SQ SEQUENCE 278 AA; 30153 MW; 368C3C505C0A2DC6 CRC64;

Query Match 2.7%; Score 8; DB 2; Length 278;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 FDVKNPGG 138
DB 123 FDVKNPGG 130

RESULT 37
Q33E52 PRELIMINARY; PRT; 278 AA.
AC Q33E52;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

```

```

DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major antigenic protein MAPI (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN SEQUENCE FROM N.A.
RP STRAIN=SOUTH EAST BOTSWANA;
RX MEDLINE=21539003; PubMed=11682561;
RA Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaïd A.,
RA Haydon D.T., van Heerden H., Allsopp B.A.;
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
RT Are Not Geographically Constrained and Show No Evidence of Having
RT Evolved under Positive Selection Pressure.";
RL J. Clin. Microbiol. 39:4200-4203 (2001).
DR EMBL; AF368015; AAK98155.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 1
FT NON_TER 278
SQ SEQUENCE 278 AA; 30156 MW; 0355DE6B85842386 CRC64;

Query Match 2.7%; Score 8; DB 2; Length 278;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 FDVKNPGG 138
DB 123 FDVKNPGG 130

RESULT 38
Q9F476 PRELIMINARY; PRT; 280 AA.
AC Q9F476;
DT 01-MAR-2001 (TrEMBLrel. 15, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P28-2 (Major outer membrane protein P30-10).
DE P28-2 OR P30-10.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN SEQUENCE FROM N.A.
RP STRAIN=JAKE;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399 (1999).
RN SEQUENCE FROM N.A.
RP STRAIN=JAKE;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis.";
RL Gene 254:245-252 (2000).
RN SEQUENCE FROM N.A.
RP STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680 (1998).
RN [4]

```

```
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF082744; AAG14357.1; -
DR EMBL; AF324792; AAK31312.1; -
DR EMBL; AF078553; AAK28696.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30362 MW; D1BB28B5AF086CA3 CRC64;

Query Match 2.7%; Score 8; DB 2; Length 280;
Best Local Similarity 100.0%; Pred.No. 9.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 281 GYFGGEIG 288
Db 268 GYFGGEIG 275
|||||

RESULT 39
ID Q9S6H0 PRELIMINARY; PRT; 281 AA.
AC Q9S6H0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Major antigenic protein 1.
GN MAP1.
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WELGEVONDEN;
RX MEDLINE=99216274; PubMed=10198207;
RA Sulsona C.R., Mahan S.M., Barbet A.F.;
RT "The map1 Gene of Cowdria ruminantium Is a Member of a Multigene
RT Family Containing Both Conserved and Variable Genes.";
RL Biochem. Biophys. Res. Commun. 257:300-305(1999).
DR EMBL; AF125274; AAD26344.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 281 AA; 30343 MW; B73D722F54A0CD75 CRC64;

Query Match 2.7%; Score 8; DB 2; Length 281;
Best Local Similarity 100.0%; Pred.No. 9.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 FDVKNPGG 138
Db 124 FDVKNPGG 131
|||||

RESULT 40
ID Q9S6H1 PRELIMINARY; PRT; 281 AA.
AC Q9S6H1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Major antigenic protein 1.
GN MAP1.
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=LEMOCOT3;
RX MEDLINE=99216274; PubMed=10198207;
RA Sulsona C.R., Mahan S.M., Barbet A.F.;
RT "The map1 Gene of Cowdria ruminantium Is a Member of a Multigene
RT Family Containing Both Conserved and Variable Genes.";
RL Biochem. Biophys. Res. Commun. 257:300-305(1999).
DR EMBL; AF125277; AAD26350.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 281 AA; 30315 MW; 9A33A1C264A438E1 CRC64;

Query Match 2.7%; Score 8; DB 2; Length 281;
Best Local Similarity 100.0%; Pred.No. 9.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 FDVKNPGG 138
Db 124 FDVKNPGG 131
|||||

Search completed: July 4, 2003, 06:44:35
Job time : 179 secs
```

~~This Page~~ Blank (uspto)